



PCT
 WELTORGANISATION FÜR GEISTIGES EIGENTUM
 Internationales Büro
 INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
 INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation ⁷ : C07H 21/00	A2	(11) Internationale Veröffentlichungsnummer: WO 00/59917 (43) Internationales Veröffentlichungsdatum: 12. Oktober 2000 (12.10.00)
(21) Internationales Aktenzeichen: PCT/EP00/02893 (22) Internationales Anmeldedatum: 31. März 2000 (31.03.00) (30) Prioritätsdaten: 199 14 808.2 31. März 1999 (31.03.99) DE (71)(72) Anmelder und Erfinder: RAUHE, Hilmar [DE/DE]; Gutenbergstrasse 29, D-45128 Essen (DE). (72) Erfinder; und (75) Erfinder/Anmelder (<i>nur für US</i>): FELDKAMP, Udo [DE/DE]; Neudorfer Strasse 141, D-47057 Duisburg (DE). BANZHAF, Wolfgang [DE/DE]; Siegenstrasse 104, D-44359 Dortmund (DE). HOWARD, Jonathan, C. [GB/DE]; Heinestrasse 19, D-50931 Köln (DE).		(81) Bestimmungsstaaten: JP, US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Veröffentlicht <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i>
(54) Title: INFORMATION-CARRYING AND INFORMATION-PROCESSING POLYMERS (54) Bezeichnung: INFORMATIONSTRAGENDE UND INFORMATIONSVERRARBEITENDE POLYMERE (57) Abstract <p>The invention relates to methods for producing information-carrying polymers, information-carrying polymers obtained using these methods, and to methods for isolating, duplicating, and selecting such information-carrying polymers. The invention also relates to polymeric data memories and DNA computers which contain the information-carrying polymers, and to the use of information-carrying polymers for producing molecular weight standards, as markers or signatures, for encoding information, as molecular-scale adhesives, or for producing or processing the smallest molecular structures.</p> (57) Zusammenfassung <p>Die Erfindung betrifft Verfahren zur Herstellung informationstragender Polymere, nach diesen Verfahren erhältliche informationstragende Polymere; Verfahren zur Isolierung, zur Vervielfältigung und zum Auslesen solcher informationstragender Polymere; polymere Datenspeicher und DNA-Computer, die informationstragende Polymere umfassen sowie die Verwendung informationstragender Polymere zur Herstellung von Molekulargewichtsstandards, als Marker oder Signaturen, zur Verschlüsselung von Information, als molekulare Kleber oder zur Herstellung oder Bearbeitung kleinster molekularer Strukturen.</p>		

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidshan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Informationstragende und informationsverarbeitende Polymere

Beschreibung

- 5 Die Erfindung betrifft Verfahren zur Herstellung informationstragender Polymere, nach diesen Verfahren erhältliche informationstragende Polymere; Verfahren zur Isolierung, zur Vervielfältigung und zum Auslesen solcher informationstragender Polymere; polymere Datenspeicher und DNA-Computer, die informationstragende Polymere umfassen sowie die Verwendung informationstragender Polymere zur Herstellung von
- 10 Molekulargewichtsstandards, als Marker oder Signaturen, zur Verschlüsselung von Information, als molekulare Kleber oder zur Herstellung oder Bearbeitung kleinster molekularer Strukturen.

- Es ist bekannt, Nukleinsäuren zur Markierung von Materialien zu verwenden. Die US-
- 15 05451505 beschreibt die Idee, Nukleinsäuren der Länge 20-1000bp zur Kennzeichnung von Materialien zu verwenden, die zum Zwecke der Authentifizierung dienen können. Ganz grundsätzliche Probleme wie das Konstruieren von geeigneten Sequenzen oder von geeigneten Kodierungen zur Repräsentation von Information werden jedoch nicht diskutiert, geschweige denn, befriedigend gelöst.

- 20 Es ist bekannt, Nukleinsäuren wie DNA (Deoxiribonucleic Acid = Desoxyribonukleinsäure) und RNA (Ribonucleic Acid = Ribonukleinsäure) auch außerhalb von Organismen zur Informationsverarbeitung zu verwenden. Es wurden bisher verschiedene Ansätze, DNA Moleküle zur Informationsverarbeitung zu verwenden, vorgestellt.

- 25 Die WO 97/07440 und [L. M. Adleman, Molecular Computation of Solutions to Combinatorial Problems, *Science*, 266, 1021-1024, (1994)] beschreiben ein Verfahren, einen Graphen-Algorithmus ("Hamiltonian-Pfad-Problem") als Entscheidungsproblem mit Hilfe von DNA Molekülen zu berechnen. Der Algorithmus wird implementiert, indem die Kanten und Knoten des Graphen als DNA Sequenzen repräsentiert werden. Die Berechnung des Algorithmus wird
- 30 dadurch vorgenommen, daß durch die Hybridisierung komplementärer DNA Sequenzen eine Menge von Pfaden des Graphen erzeugt werden.

- Aus dieser Menge von Pfaden werden dann mit Hilfe molekularbiologischer Verfahren alle Pfade entfernt, die eine falsche oder keine Lösung des Algorithmus sind. Wurden alle Schritte erfolgreich ausgeführt, so muß am Ende entweder mindestens ein DNA Strang übrigbleiben,
- 35 der die richtige Lösung des Problems enthält, oder es bleibt kein DNA Strang übrig, was bedeutet, daß der Algorithmus keine Lösung hat.

- Das beschriebene Verfahren setzt voraus, daß genügend viele Moleküle zur Verfügung stehen, damit jede mögliche Lösung der jeweiligen Probleminstanz statistisch mindestens einmal in der Ausgangsmenge von Pfaden vorkommt. Dies kann aufgrund der statistischen
- 40 Natur der Hybridisierungsvorgänge jedoch nicht garantiert werden. Außerdem können im

BESTATIGUNGSKOPIE

ersten Schritt auch zyklische Graphen entstehen, was von vorneherein die Menge falscher Lösungen vergrößert.

Der in WO 97/07440 beschriebene Ansatz zeigt, daß Symbolverarbeitung überhaupt in vitro mit DNA Molekülen vorgenommen werden kann. Jedoch ist der beschriebene Ansatz in der

- 5 Praxis kaum verwendbar: Ein Problem ist, daß der Algorithmus nicht deterministisch, sondern nur stochastisch ist, das gefundene Ergebnis ist damit nur mit einer bestimmten Wahrscheinlichkeit korrekt. Ein weiteres Problem ist, daß die Implementierung des Algorithmus nicht effizient (Laufzeit-optimal) ist, dadurch wird die Berechnung langsam.

- Es wird dargelegt, daß das verwendete Verfahren zur Lösung NP-vollständiger Probleme
10 (einer Klasse von Problemen, für die nur deterministische Lösungsalgorithmen mit exponentieller Laufzeit bekannt sind und von der vermutet wird, daß es keine effizienten deterministischen Lösungsalgorithmen gibt) besonders gut geeignet sei, weil der Algorithmus aufgrund der Parallelität der ausgeführten Operationen nur lineare Laufzeit benötige. Diese Argumentation ist jedoch insofern fehlerhaft, als die lineare Laufzeit durch eine exponentielle
15 Anzahl von Molekülen kompensiert wird. Die exponentielle Problemgröße bleibt also unverändert bestehen.

- Insgesamt ist das in der WO 97/07440 beschriebene Verfahren zu sehr eingeschränkt, um für molekulare Informationsverarbeitung über den beschriebenen Algorithmus hinaus verwendbar zu sein: Das Hamiltonian-Pfad-Problem ist fest kodiert, andere Algorithmen können damit
20 nicht berechnet werden, jede Problemistanz muß neu kodiert werden. Programmierung ist nicht vorgesehen, alle Schritte des Algorithmus werden von Hand ausgeführt. Ein Input-/Output-System ist nicht vorgesehen. Insgesamt ist das Verfahren zur Programmierung von Algorithmen oder zur Implementierung eines Computers nicht geeignet.

- Die WO 97/29117 und [Frank Guarnieri, Makiko Fliss, Carter Bancroft, Making DNA Add,
25 *Science*, 273, 220-223, (1996)] beschreiben ein Verfahren, Additionen mit Hilfe von DNA Molekülen auszuführen. Die Addition erfolgt als Schiebeoperation mit Überlauf-Übertrag in vitro mit DNA Molekülen und Primer-Elongation.

- Das beschriebene Verfahren beschreibt ausschließlich Additionen. Jedoch ist selbst die Verwendung als Addierer aus mindestens zwei Gründen ungünstig: Zum einen ist Addition mit
30 dem beschriebenen Verfahren nicht effizient (Laufzeit-optimal) implementiert. Zum anderen ist das verwendete System formal unvollständig, weil die durch Addition erzeugten Ergebnisse ihrerseits nicht mehr als Zahlen für Berechnungen verwendet werden können. Über die Addition hinausgehende Konzepte fehlen. Insgesamt ist das Verfahren zur Programmierung von Algorithmen oder zur Implementierung eines Computers nicht geeignet.

- 35 In [Qi Ouyang, Peter D. Kaplan, Shumao Liu, Albert Libchaber, DNA Solution of the Maximal Clique Problem, *Science*, 278, 446-449, (1997)] wird ein Verfahren beschrieben, einen Algorithmus zur Lösung des "Max-Clique-Problem"s zu implementieren. Der Algorithmus stammt aus der Graphentheorie und fällt unter die NP-vollständigen Probleme. Wie in [L. M. Adleman, Molecular Computation of Solutions to Combinatorial Problems, *Science*, 266,
40 1021-1024, (1994)] ist das Verfahren nur in der Lage, einen fest kodierten Lösungsalgorithmus zu berechnen. Auch hier wird das Ergebnis nur mit einer bestimmten

Wahrscheinlichkeit erhalten. Das von den Autoren beschriebene Verfahren enthält keine weiterführenden Konzepte (etwa in Hinsicht der Implementierung auch anderer Algorithmen). Insgesamt ist das Verfahren zur Programmierung von Algorithmen oder zur Implementierung eines Computers nicht geeignet.

- 5 In [Erik Winfree, Xiaoping Yang, Nadrian C. Seeman, Universal Computation via Self-assembly of DNA: Some Theory and Experiments, *Proceedings of the 2nd DIMACS Meeting on DNA Based Computers, Princeton University, June 20-12, (1996)*] wird erörtert, reguläre Grammatiken in DNA durch Verknüpfung von Oligonukleotiden mit "sticky ends" zu implementieren. Die Ausführungen sind jedoch rein theoretischer Natur und scheitern an wesentlichen, bisher ungelösten Problemen. Insbesondere wird nicht dargelegt, wie die zur Implementierung von Grammatiken benötigten Sequenzen konstruiert werden müssen. Gerade dies ist aber das entscheidende Problem, ohne dessen Lösung Grammatiken nicht implementiert werden können. Der Grund dafür liegt darin, daß die Sequenzen, die die Variablen und Terminale einer Grammatik repräsentieren, sowohl eindeutig, als auch untereinander hinreichend unähnlich sein müssen. Außerdem müssen sie bestimmte strukturelle, chemische und physikalische Eigenschaften erfüllen. Andernfalls sind Fehlhybridisierungen zwischen Sequenzen zwangsläufig, die bei der Polymerisierung zu ungewollten Kettenverlängerungen und Kettenabbrüchen führen und damit das Funktionieren des Verfahrens unmöglich machen. Dieses Problem verschärft sich mit der Anzahl der benötigten Sequenzen exponentiell.

- Das beschriebene Verfahren wird von den Autoren selbst wieder verworfen, bzw. insofern als unzureichend bezeichnet, als es keine sonderlich interessanten Berechnungen erlaube (Erik Winfree, Xiaoping Yang, Nadrian C. Seeman, Universal Computation via Self-assembly of DNA: Some Theory and Experiments, *Proceedings of the 2nd DIMACS Meeting on DNA Based Computers, Princeton University, June 20-12, (1996)*, S. 8, 2. Absatz, 1. Zeile).

- Weiterführende Experimente der Autoren stützen sich entsprechend auch nicht weiter auf reguläre Grammatiken und lineare Polymere, sondern auf kontextsensitive Grammatiken zur Erzeugung von DNA Gitterstrukturen [Erik Winfree, Furong Liu, Lisa A. Wenzler & Nadrian C. Seeman, Design and self-assembly of two-dimensional DNA crystals, *Nature*, **394**, 539-544, (1998)]. Der von den Autoren beschriebene Ansatz ist womöglich zur Erzeugung DNA-basierter Nanostrukturen (Herstellung von Katalysatoren usw.) geeignet. Zur Informationsverarbeitung eignet er sich hingegen nicht, hier sind die erzeugten Gitterstrukturen eher hinderlich.

- Zwar sprechen die Autoren von der Möglichkeit, das mathematische Konzept der "Wang tiles" durch die erzeugten Gitterstrukturen zu realisieren und damit potentiell auch für Berechnungen zu verwenden, sie bleiben aber bei der bloßen Erwähnung einer solchen Möglichkeit, ohne zu beschreiben, wie dies im Sinne einer Informationsverarbeitung genutzt werden kann.

- Es ist überhaupt zweifelhaft, ob der von den Autoren beschriebene Ansatz überhaupt zu einer Informationsverarbeitung geeignet ist: Die Gitterstrukturen verhindern die Vervielfältigung informationstragender Sequenzen (z.B. durch Klonierung oder PCR = Polymerase Chain

Reaction = Polymerase Kettenreaktion) und machen es unmöglich, die erzeugten Strukturen als Daten zu verwenden, weil diese nicht mehr, etwa durch PCR, ausgelesen werden können. Entsprechend sieht der gewählte Ansatz in der jetzigen Form konzeptionell gar keine Möglichkeit vor, die erzeugten Strukturen im Sinne einer Informationsverarbeitung, etwa als Daten, zu nutzen.

Aus dem Stand der Technik ist bisher kein Verfahren bekannt, das es erlaubt, DNA Moleküle für die Implementierung effizienter Algorithmen zu verwenden. Es ist kein Verfahren bekannt, Berechnungen automatisch (etwa in Art eines molekularen Computers) durchzuführen. Es ist auch kein Verfahren bekannt, Programme in Form von regulären Grammatiken mit molekularen Verfahren so zu implementieren, daß damit

- a) unterschiedliche (im Idealfall beliebige) reguläre Grammatiken implementiert werden können
- b) die mit regulären Grammatiken erzeugten Wörter einer Sprache ausgelesen werden können
- c) die mit regulären Grammatiken erzeugten Wörter einer Sprache für technische Zwecke (z.B. Informationsverarbeitung, Polymerchemie) weiterverwendet werden können.

Der vorliegenden Erfindung liegt daher die Aufgabe zugrunde, ein Verfahren bereitzustellen, das es erlaubt, Informationsverarbeitung auf molekularer Basis mithilfe regulärer Grammatiken vorzunehmen, ohne dabei auf eine oder wenige Grammatiken eingeschränkt zu sein. Das Verfahren soll kompatibel zu herkömmlichen Computern sein und eine Informationsverarbeitung erlauben, die teilweise auf molekularer Basis und teilweise auf der Basis herkömmlicher Computer stattfindet. Das Verfahren soll programmierbar und weitgehend automatisierbar sein. Die innerhalb des Verfahrens erzeugten Polymere sollen lesbar und für weiterführende Anwendungen und Verfahren verwendbar sein.

Diese Aufgabe wird gelöst durch ein Verfahren zur Herstellung informationstragender Polymere, das umfaßt,

- I. eine reguläre Grammatik $G = (\Sigma, V, R, S)$ mit einem endlichen Terminalalphabet Σ , einer endlichen Variablenmenge V , einer endlichen Regelmenge R und einem Startsymbol S zu definieren;
- II. das NFR-Verfahren (Niehaus-Feldkamp-Rauhe-Verfahren) zur Herstellung von Monomersequenzen (Oligo- oder Polymere);
- III. mit dem NFR-Verfahren eine in Schritt I definierte Grammatik zu implementieren, indem damit Monomersequenzen hergestellt werden, die die Regelmenge R einer Grammatik G eindeutig darstellen;
- IV. aus den in Schritt III hergestellten Monomersequenzen für jede Regel der Regelmenge R von G ein die Regel repräsentierendes Oligomer (*Algomer*) zusammenzusetzen (*Algomer-Assemblierung*);

- V. die in Schritt IV zusammengesetzten Oligomere (*Algomere*) zu informationstragenden Polymeren zu verknüpfen (*Symbolpolymerisation*).

Im Rahmen der vorliegenden Erfindung werden folgende Definitionen und Abkürzungen verwendet:

Algomer	Doppelsträngiges Oligomer, das eine Regel einer gegebenen Grammatik repräsentiert. Algomere können miteinander zu Logomeren verknüpft werden.
10 Auslese-PCR	PCR, die zum Auslesen der in Logomeren enthaltenen Information verwendet wird.
Biochip	Auch als Microarray, DNA-Array, Gene-Array, Gene-Chip bezeichneter Träger einer Anzahl von Nukleinsäuren, die zur Detektion komplementärer Sequenzen benutzt werden.
15 Bitpolymerisation	Prozeß des Verkettens von Algomeren zu Logomeren, wenn es nur zwei verschiedene Elongatoren gibt.
Byte	Informationseinheit aus 8 Bit, bzw. Molekül, das 8 Bit repräsentiert.
Elongator	Algomer mit zwei Überhangsequenzen; kann mit Terminator oder Elongator ligieren und führt zu einer Kettenverlängerung.
20 Grammatik	Formalismus, der Sprachen beschreibt. Er basiert auf der formalen Theorie von Sprachen [Chomsky, N., Three models for the description of language, <i>JACM</i> , 2:3, 113-124, (1956)], [Chomsky, N., On certain formal properties of grammars, <i>Inf. and Control</i> , 2:2, 137-167, (1959)], [Chomsky, N., Formal properties of grammars, <i>Handbook of Math. Psych.</i> , 2, 323-418, (1963)].
25	Eine Grammatik G beschreibt eine Sprache L(G), das Alphabet dieser Sprache und deren Syntax. Nach einer Grammatik G können alle Wörter dieser Sprache erzeugt werden.
30	Eine Grammatik G ist ein Quadrupel (Σ , V, R, S) mit Terminalalphabet Σ , Variablenmenge V, Startsymbol S und Regelmenge R.
Logomer	Polymer, das symbolische Informationen trägt und durch die Verkettung von Algomeren erzeugt wurde. Entsprechend besteht ein Logomer aus sich wiederholenden Einheiten von Algomeren. Ein Logomer repräsentiert ein Wort einer Sprache L(G), die von der entsprechenden Grammatik G erzeugt wird.
35 Monomer	Einzelmolekül. Mehrere Monomere können zu längeren Ketten verknüpft werden und so Oligomere und Polymere bilden.
40	Monomere sind im Fall von Desoxyribonukleinsäure die Nukleotide (Adenin, Cytosin, Guanin, Thymin wie auch sog. Basenanaloga wie Hypoxanthin etc.).
Multibyte	Eine beliebige Datenstruktur, die aus Vielfachen von Bytes besteht.

	Oligomer	Kurzkettiges Molekül aus sich wiederholenden Einheiten (Monomeren). Auch kurze doppelsträngige Moleküle werden als Oligomere bezeichnet.
	PCR	Polymerase Chain Reaction = Polymerase Kettenreaktion; Verfahren zur exponentiellen Vervielfältigung von DNA.
5		Die PCR benötigt ein DNA Template, das vervielfältigt wird, und zwei Primer, die jeweils gegenläufig im Template ansetzen und als Startpunkte einer DNA Polymerisation fungieren. Durch iterative Wiederholung von Schmelzen-Hybridisierung-Polymerisations-Zyklen wird das DNA Template vervielfältigt.
10	Polymer Regel	Langkettiges Molekül aus sich wiederholenden Einheiten (Monomeren). auch: Produktionsregel, Ersetzungsregel oder Ableitungsregel. Beschreibt die Ersetzung von Symbolen durch andere. Durch die wiederholte Anwendung von Regeln können Symbolketten erzeugt werden.
15	Sequenz	Informatik: Abfolge von Zeichen; Chemie: Abfolge kovalent verbundener Monomere; Molekularbiologie: Abfolge kovalent verbundener Nukleotide.
	Komplementarität	Zwei Sequenzen sind dann komplementär, wenn sie miteinander hybridisieren können. Im Falle von DNA sind z.B. die Sequenzen 5'attt3' und 5'aaat3' komplementär, die Sequenz 5'acgt3' ist zu sich selbst komplementär.
20		
	Startsymbol	Variable innerhalb einer Grammatik, von der ausgehend, durch Anwendung der Regeln der Grammatik, eine Symbolkette erzeugt werden kann.
25	Symbolpolymerisation Terminal	Prozeß des Verkettens von Algomeren zu Logomeren. Symbol einer Grammatik. Ein Terminal kann nicht weiter ersetzt (substituiert) werden. Terminale sind die "Buchstaben" der Worte einer Sprache L(G).
	Terminator	Algomer mit einer Überhangsequenz. Kann nur mit Elongator ligieren. Führt zum Kettenabbruch bei einer Symbolpolymerisation.
30		
	Uniqueness	Eindeutigkeit von Sequenzen untereinander. Eine Sequenz S aus Monomeren ist 10-unique zu einer Menge M anderer Sequenzen, wenn jede Teilsequenz der Länge 10 von S in keiner anderen Sequenz der Menge M auftritt. Die Uniqueness kann auch in Prozent angegeben werden. Z.B. sind 2 Sequenzen der Länge 20, deren längste gemeinsame Teilsequenz 5 Monomere beträgt, zueinander 6-unique, ihre Uniqueness in Prozent beträgt: $(1 - 5/20) * 100 = 80\%$.
35		
	Variable	Symbol einer Grammatik. Eine Variable kann gemäß der Regel einer Grammatik von Terminalen, Variablen oder Kombinationen von Terminalen und Variablen ersetzt werden.
40		

Kurzbeschreibung der Abbildungen

Abbildung 1: Algomere, wie in der Erläuterung zu Schritt IV des erfindungsgemäßen Verfahrens beschrieben. Die Algomere besitzen jeweils eine doppelsträngige Kernsequenz, die ein Terminal einer gegebenen Grammatik darstellt und mindestens eine einzelsträngige Überhangsequenz, die eine Variable der gegebenen Grammatik darstellt, so daß jeweils ein Algomer genau eine Regel der gegebenen Grammatik repräsentiert. X und Y sind keine Variablen, sondern Überhangsequenzen, die z.B. für eine Klonierung benutzt werden können. Mit Überstrichen dargestellte Buchstaben kennzeichnen Komplementarität. Gemäß der Definition sind A0A und A1A Elongatoren, XsA und AeY Terminatoren. Die hier dargestellten Algomere repräsentieren die im Folgenden beschriebenen Grammatik zur Erzeugung binärer Zufallszahlen.

Abbildung 2: Symbolpolymerisation, wie in der Erläuterung zum erfindungsgemäßen Schritt V beschrieben. Algomere werden durch Verkettung (im Falle von DNA: Hybridisierung und Ligation) zu Logomeren verknüpft. Das Logomer Xs01010101eY beinhaltet eine terminierte Bitfolge, die durch die Überhänge X und Y in einer definierten Orientierung vervielfältigt werden kann.

Abbildung 3: Muster, das die Symbolpolymerisation für binäre Zufallszahlen nach Gelelektrophorese und Färbung zeigt (Spuren 1-3). Aufgrund der zufälligen Länge der binären Zufallszahlen ergibt sich ein regelmäßiges Leitemuster. Spur 4 zeigt einen 50bp Molekulargewichtsstandard (Gibco BRL, Life Technologies, Katalog Nr. 10416-014).

Abbildung 4: Klonierung eines durch Symbolpolymerisation erhaltenen Logomers in einem Vektor, wie zu Vervielfältigung und Isolierung von Logomeren im Folgenden beschrieben.

Abbildung 5: Schematische Darstellung eines Bandenmusters, das durch Auslesen eines binären Logomers durch PCR und nachfolgende Gelelektrophorese erhalten wird (im Folgenden beschrieben). Im gezeigten Beispiel haben die Algomere eine Länge von 30bp (Überhangsequenzen nur 1/2 gerechnet). Bei bekannter Länge des Logomers reicht es, jeweils nur 0en oder nur 1en auszulesen. Beide Bits auszulesen dient der Kontrolle. Das Verfahren kann auch für mehrwertige Logomere (mehr als 2 Elongatoren) verwendet werden.

Abbildung 6: Bandenmuster einer Gelelektrophorese nach PCR zum Auslesen von drei verschiedenen Logomeren. Die Logomere wurden nach der im Folgenden beschriebenen Grammatik für Zufallszahlen beliebiger Länge erhalten. Als Zufallszahlen von unten nach oben gelesen ist a = 262, b = 97, c = 329. M (Spur 5) ist ein 50bp Molekulargewichtsstandard (Gibco BRL, Life Technologies, Katalog Nr. 10416-014).

Abbildung 7: Schematische Darstellung des Auslesens von Logomeren durch Restriktionsverdau wie im Folgenden beschrieben. Die Elongatoren tragen Restriktionsschnittstellen, die derart asymmetrisch angeordnet sind, daß der Restriktionsverdau von Logomeren in einem eindeutigen Schnittmuster von Fragmenten resultiert. Das Schnittmuster entspricht Banden verschiedener Länge, und kann durch Gelelektrophorese sichtbar gemacht werden. R1 und R2 sind unterschiedliche Restriktionsenzyme, x und y bezeichnen die Länge der nach Restriktionsverdau erhaltenen Fragmente. Zweckmäßig ist z.B. ein Verhältnis x:y von 1:2.

10 Abbildung 8: Bandenmuster, die durch Gelelektrophorese nach dem Auslesen von Logomeren durch Restriktionsverdau erhalten werden. Spur 7 und 8 enthalten Molekulargewichtsstandards (Spur 7: 50pb Leiter, Gibco BRL, Life Technologies, Katalog Nr. 10416-014; Spur 8: 10bp Leiter)

15 Abbildung 9: Schematische Darstellung eines polymeren Datenspeichers auf der Basis von Algomeren und Symbolpolymerisation wie im Nachfolgenden beschrieben. Algomere können an Ankermoleküle (AM), die an einem festen Träger (C) gebunden ist, polymerisieren. Das Schreiben (W) erfolgt durch die wiederholte Abfolge (ReW) von Hybridisierungs-Ligations-Restriktionszyklen (Hyb, Lig, Res). Durch Denaturieren oder Restriktion (Den/Dig) können die
20 erhaltenen Logomere abgetrennt und ausgelesen werden.

Abbildung 10: Vereinfachte Darstellung der Bestandteile eines DNA Desktop Computers wie im Nachfolgenden beschrieben. Im Einzelnen sind:

A: Oligo-Synthesizer, B: Thermozykler, C: Reaktionskammern, D: Pipettiervorrichtung, E: Gel,
25 F: Scanner, G: Steuerungscomputer

Es kennzeichnen:

1: Zugabe von Oligonukleotiden, 2: Zugabe synthetisierter Oligomere in Reaktionskammern des Thermozyklers, 3: Zugabe von Lösungen und Molekülen, die zur Algomer-Assemblierung, Symbolpolymerisation, zur Isolierung von Logomeren und zum Auslesen benötigt werden.

30 Abbildung 11: Verschlüsselung von Logomeren wie im Folgenden beschrieben. Sind die Terminatoren und die darin primenden Primer unbekannt, so ist das jeweilige Logomer verschlüsselt und kann nicht ausgelesen werden (A). Ist dagegen ein in einem Terminator primender Primer verfügbar oder die Sequenz eines Terminators bekannt, so kann das
35 jeweilige Logomer gelesen werden (B).

Abbildung 12: Y-förmiges Molekül, das als Terminator bei der asymmetrischen Verschlüsselung von Logomeren verwendet werden kann. Elongatoren können an das mit 2 markierte Ende angeknüpft werden, weitere, z.B. Y-förmige, Moleküle an die mit 1 und 3
40 bezeichneten Enden. Werden mehrere Y-förmige Moleküle miteinander verknüpft, so erhält man baumartige Strukturen.

Abbildung 13: Logomer mit Terminatoren s und e, die als baumartige Strukturen realisiert sind.

- 5 Abbildung 14: Mit einem Vektor V verknüpft Logomer L mit baumartigen Terminatoren. Die Terminatoren sind so konstruiert, daß jeweils nur ein Ast des Terminators mit dem Vektor verknüpft werden kann.

- 10 Abbildung 15: Markierung von Nukleinsäuren mit Logomeren wie im Nachfolgenden beschrieben: Um gentechnisch hergestellte oder veränderte Produkte zu markieren, wird ein Logomer mithilfe rekombinativer Techniken in einen nicht-transkribierten Bereich, z.B. vor den Promotor (P) eines zu markierenden Genes (G), eingesetzt.

- 15 Abbildung 16: Markierung von Dokumenten mit Logomeren wie im Nachfolgenden beschrieben. Spur 1 zeigt den verwendeten Molekulargewichtsstandard (50bp, Gibco BRL, Life Technologies, Katalog Nr. 10416-014), Spur 2 (0-Bits) und Spur 3 (1-Bits) das Bandenmuster des Auslesens von Logomer Nr. 330 durch PCR aus wässriger Lösung, die Spuren 4 und 5 dasselbe wie die Spuren 2 und 3, wobei hierbei jedoch das Logomer Nr. 330 nicht aus wässriger Lösung, sondern in ca. 10^9 Molekülen/ μ l getrocknet auf Papier (3M Post-It, 1 Stunde getrocknet) als Papierschnipsel (ca. 1mm^2) in die Auslese-PCR einging.

- 25 Abbildung 17: Herstellung von Molekulargewichtsstandards durch Auslese-PCR, die ein unäres Logomer als Template enthält, wie im Nachfolgenden beschrieben. 1 zeigt die Anordnung verschachtelter Primer in der Auslese-PCR, 2 das nach Gelelektrophorese der PCR-Fragmente erhaltene Bandenmuster.

- 30 Abbildung 18: Verkleben von Flächen mit Nukleinsäuren wie im Nachfolgenden beschrieben. C bezeichnet die zu verklebende Fläche, Log die Logomere, die zum Verkleben an die Flächen gebunden werden. 1 zeigt das Verhalten von Flächen mit nicht-komplementären, 2 das Verhalten von Flächen mit komplementären Logomeren.

- 35 Abbildung 19: Verkleben von Flächen mit Logomeren, Antikörpern und Liganden wie im Nachfolgenden beschrieben: C₀ und C₁ bezeichnen die zu verklebenden Flächen, die aus gleichem oder unterschiedlichem Material bestehen können. Die zu verklebenden Flächen sind mit Logomeren (Log) versetzt. An die Logomere können Proteine, z.B. Antikörper vom Typ Ab A und Ab B binden, wobei Ab A und Ab B gleich oder verschieden sein können. Ab A und Ab B sind ihrerseits durch einen Liganden (Li) miteinander verbunden.

- 40 Abbildung 20: Verkleben von Flächen mit Proteinen z.B. Antikörpern ohne Verwendung von Logomeren. C₀ und C₁ bezeichnen die zu verklebenden Flächen, die aus gleichem oder unterschiedlichem Material bestehen können. Die Proteine vom Typ Ab A und Ab B binden

direkt an die zu verklebenden Flächen und sind ihrerseits durch einen Liganden (Li) miteinander verbunden.

Abbildung 21: Bei der Verknüpfung von Sequenzen zu längeren Sequenzketten kann es zu Verletzungen der vorgegebenen Uniqueness kommen, die Fehlhybridisierungen und damit Disfunktionalität nach sich ziehen kann. Dies ist auf das Entstehen neuer Basissequenzen (in der Abbildung markiert) aufgrund der Verknüpfung zurückzuführen.

Abbildung 22: Beispiel des Verknüpfens von Terminalen mit einer Variablen. Durch die gezeigten vier Regeln einer Grammatik mit der Variable A ergeben sich vier verschiedene Pfade, die sich über die Länge der Variablensequenz für A überschneiden. Außerdem überschneiden sich auch einige Sequenzen für mehrfach vorkommende Terminale (b, c), so daß links und rechts je drei Pfade zusammenlaufen.

Abbildung 23: Haben mehr als vier verschiedene Variablensequenzen (A, B, C, D, E) Übergänge zu der gleichen Terminalsequenz (0), so muß mindestens eine Basissequenz mehrfach verwendet werden. Der gepunktete Rahmen zeigt den Iterationsschritt an, in der eine Verletzung der Uniqueness toleriert werden muß, um alle Regeln aus R übersetzen zu können.

Abbildung 24: Paralleles Auffüllen zweier Sequenzenverlängerungen für die Variablen A und B. Der gepunktete Rahmen zeigt den Iterationsschritt, in dem die Startbasissequenzen für die Pfadsuche liegen. Ab dem nächsten Iterationsschritt können ggf. Verletzungen der Uniqueness toleriert werden. Zu beachten ist, daß sich hier Verzweigungen auch über gruppengrenzen hinweg ergeben (z.B. des Terminal b zu den Variablen A und B).

Abbildung 25: Skizze eines 1-Byte repräsentierenden Moleküls. Der mit x markierte Abschnitt enthält eine eindeutige Sequenz, die einen definierten Bytewert darstellt (die für die Darstellung aller Bytewerte benötigten 256 Nukleinsäuresequenzen sind im Sequenzprotokoll aufgeführt). Der mit s markierte Abschnitt enthält eine eindeutige Sequenz, die die Byteposition des betreffenden Bytes innerhalb von Multibytes kodiert und dient als Template für PCR Reaktionen (siehe Sequenzprotokoll). Die mit o und e markierten Abschnitte dienen zur Herstellung von Multibytes aus einzelnen Bytes und als Template für PCR Reaktionen (siehe Sequenzprotokoll).

Abbildung 26: Schematische Darstellung von vier 1-Byte Molekülen mit unterschiedlichen Bytepositionen ($s_0 - s_3$). Die Einzelbytes können zu Multibytes verbunden werden.

Abbildung 27: Zum Zwecke der Vervielfältigung können einzelne Bytemoleküle in genetischen Vektoren (Plasmiden) kloniert werden.

Abbildung 28: Beispiel der Konstruktion eines Byte-repräsentierenden DNA Moleküls. Das Moleküle besteht aus mehreren funktionellen Untereinheiten: Für X gibt es 256 eindeutige Basenfolgen die alle Werte eines einzelnen Bytes darstellen. S ist eine eindeutige Sequenz, die die Position eines Bytes (das im Strang unmittelbar folgende X) innerhalb von Multibytes repräsentiert. O und E dienen als eindeutige Erkennungssequenzen für die Verkettung einzelner Bytes zu Multibytes.

DNA Bytes können unverkettet oder verkettet zur Kennzeichnung verwendet werden. Durch das Ausschneiden der X oder der SX Untereinheiten werden DNA Abschnitte gewonnen, die für das "Spotten" von Biochips verwendet werden. Die so hergestellten Biochips werden ihrerseits zum Auslesen von einzelnen Bytes oder Multibytes verwendet.

Abbildung 29: Verkettung von Bytes zu Multibytes. Im Beispiel werden 4 Bytes zu einer 32-bit Datenstruktur verknüpft. Außerdem können die endständigen Sequenzen L und R als Adapter fungieren, um die 32-bit DNA Datenstruktur in andere DNA einzubringen. Sie können entweder spezifische Rekombinationssites oder Restriktionssites tragen. Ein Beispiel ist die Kennzeichnung eines Plasmids in Abbildung 31.

Abbildung 30: Schematischer Aufbau eines 32-Bit Moleküls, das durch Verknüpfung von 4 1-Byte Molekülen hergestellt wird. Das Molekül kann zum Zwecke der Kennzeichnung den zu markierenden Substanzen beigemischt oder angeheftet werden. Zur Kennzeichnung von Nukleinsäurekonstrukten und Genen können die mit L und R bezeichneten Sequenzen Restriktions- oder Rekombinationssites tragen, durch die sie mit dem zu markierenden Molekül verbunden werden.

Abbildung 31: Kennzeichnung eines Plasmids durch ein 32-bit Molekül. Das nullte Byte hat den Wert 109, das erste den Wert 67, das zweite den Wert 35, das dritte den Wert 192. Als 32-Bit Zahl ohne Vorzeichen entspricht das Bytemuster des gekennzeichneten Plasmids der Zahl 3223536493.

Abbildung 32: Skizze eines 1-Byte Biochips, der mit allen 256 x-Fragmenten (siehe Abbildung 25 bis Abbildung 28) von x_0 bis x_{255} gespottet wurde (X-Chip). Dargestellt sind nur einige der 256 unterschiedlichen Sequenzen. Zum Auslesen von Multibytes werden die jeweiligen Einzelbytes mit PCR voramplifiziert (s bis e) und einzeln auf getrennten Chips hybridisiert. So sind für ein 4-Byte Molekül (siehe Abbildung 30) 4 PCR Reaktionen und 4 Chips erforderlich.

Abbildung 33: Skizze eines 1-Byte Biochips, der mit allen 256 s_0x -Fragmenten (siehe Abbildung 25 bis Abbildung 28) von s_0x_0 bis s_0x_{255} gespottet wurde (SX-Chip). Dargestellt sind nur einige der 256 unterschiedlichen Sequenzen. Im Gegensatz zum Chiptyp aus Abbildung 32 können hier die Hybridisierungsbedingungen so gewählt werden, daß Bytes abhängig von ihrer Position in einem Multibyte detektiert werden. Im Beispiel werden nur s_0x_i aber keine s_nx_i mit $n \neq 0$ detektiert. Analog lassen sich 1-Byte Chips herstellen, die nur s_1x_i detektieren usw..

Das ermöglicht es, Chips herzustellen, die Multibytes direkt (ohne PCR) und vollständig lesen können. Ein Beispiel ist der in Abbildung 37 gezeigte 4-Byte Chip.

Abbildung 34: Layout des 1-Byte Chips. Ist der Chip als X-Chip implementiert (siehe Abbildung 32) enthält er 256 Spots mit allen 256 x-Fragmente (siehe Sequenzprotokoll), ist er als SX-Chip (siehe Abbildung 33) implementiert enthält er 256 Spots mit allen s_x -Fragmenten. Jedes x-Fragment repräsentiert genau einen Bytewert ($x_0 = 0, x_1 = 1, \dots, x_{255} = 255$). Die Sequenzen für s und x sind so gewählt, daß sie untereinander möglichst ähnliche Schmelztemperaturen haben, dabei aber möglichst unterschiedliche Sequenzen, um Fehlhybridisierungen auszuschließen.

Abbildung 35: Herstellung von Multibyte-Chips aus 1-Byte SX-Chips. Im Beispiel ist die Herstellung eines 4-Byte Chips aus je einem SX_0 -, SX_1 -, einem SX_2 - und einem SX_3 -Chip gezeigt. Multibyte-Chips können z.B. zu linearen (a) oder 2-dimensionalen (b) Byte-Arrays angeordnet werden.

Abbildung 36: Auslesen eines 32-Bit Moleküls (siehe Abbildung 30 und Abbildung 31) mit einem 4-Byte SX-Chip (siehe Abbildung 35). Zum Auslesen können 32-Bit Moleküle direkt mit dem gesamten Chip hybridisiert werden. Dadurch kann der 32-bit Wert direkt ausgelesen werden (im Beispiel markiert). Außerdem können die Bytes auch unabhängig voneinander mit PCR amplifiziert und getrennt voneinander hybridisiert werden.

Alternativ dazu kann ein Multibyte-Chip auch aus identischen 8-Bit Einheiten aufgebaut sein, die nur mit X-Fragmenten gespottet sind. Um die Positionsinformation der Bytes beizubehalten müssen dann die einzelnen Bytes getrennt voneinander in den jeweils korrespondierenden Sektoren gespottet werden.

Abbildung 37: Multibyte Array aus identischen 1-Byte Chips (X-Chips). Multibyte Arrays aus X-Chips können ebenfalls zum Auslesen von Markierungsmolekülen verwendet werden wie bereits beschrieben wurde. Außerdem können Multibyte Arrays zum Speichern und für die optische Anzeige von Computerdaten verwendet werden.

Sequenzprotokoll: Moleküle, die für die Herstellung der unten beschriebenen 32-Bit Moleküle benötigt werden. Die Moleküle werden wie unten beschrieben zu Algotemen zusammengesetzt. Die o, s, e und x Einheiten werden wie unten beschrieben zu Bytes zusammengesetzt. Diese werden wiederum wie unten beschrieben zu Multibytes (im Beispiel zu 4-Byte = 32-Bit Molekülen) zusammengesetzt. Die 32-Bit Moleküle werden ihrerseits z.B. zur Kennzeichnung und Markierung verwendet und außerdem zur Herstellung von Biochips (wie unten beschrieben) verwendet.

Erläuterung zu I: Wahl von Grammatiken

Eine Grammatik ist ein Quadrupel $G = (\Sigma, V, R, S)$ mit einem endlichen Terminalalphabet Σ , einer endlichen Variablenmenge V , einer endlichen Regelmenge R und einem Startsymbol S .

- 5 Für eine Sprache $L(G)$, die durch G beschrieben wird, können alle Wörter aus L als Wörter über dem Terminalalphabet Σ gebildet werden, indem man das Startsymbol S entsprechend den Regeln aus R ableitet.

- Die Definition einer Grammatik G nach Schritt I des erfindungsgemäßen Verfahrens ist insofern frei, als beliebige, endliche Mengen von Terminalen und Variablen kodiert werden können. Erfindungsgemäß bevorzugt ist jedoch die Definition von Grammatiken, die die
- 10 Binärdarstellung von Daten gestattet, insbesondere von Grammatiken mit

$\Sigma := \{0, 1, s_0, s_1, s_2, \dots, s_{n-1}, s_n, e_0, e_1, e_2, \dots, e_{m-1}, e_m\}$ mit $n, m \in \mathbb{N}$ und $n, m \geq 0$.

Diese ermöglichen die Herstellung binärer Logomere.

- Zur Veranschaulichung wird im Folgenden eine erfindungsgemäß definierte reguläre
- 15 Grammatik zur Erzeugung von Zufallszahlen beliebiger endlicher Länge in Binärdarstellung gezeigt:

Es sei eine Grammatik $G = (\Sigma, V, R, S)$ mit Terminalalphabet $\Sigma := \{0, 1, s, e\}$, Variablenmenge $V := \{A\}$, Startsymbol S und Regelmenge

$R :=$

- 20 {
 $S := sA$
 $A \rightarrow 0A$
 $A \rightarrow 1A$
 $A \rightarrow e$
- 25 },
wobei
 $s := \text{Start}$
 $e := \text{Ende}$
sind.

30

Mit dieser Grammatik können Wörter über dem Terminalalphabet

$\Sigma := \{0, 1, s, e\}$

gebildet werden. Alle Wörter der Sprache L beginnen mit „s“ und hören mit „e“ auf, dazwischen befindet sich eine zufällige Anzahl (auch 0) zufälliger Bits („0“ und „1“).

- 35 Wörter aus der Sprache L , die nach R gebildet werden, sind beispielsweise:

- a) $S \rightarrow sA \rightarrow se$
b) $S \rightarrow sA \rightarrow s1A \rightarrow s1e$
c) $S \rightarrow sA \rightarrow s0A \rightarrow s00A \rightarrow s001A \rightarrow s0010A \rightarrow s0010e$
d) $S \rightarrow sA \rightarrow s1A \rightarrow s10A \rightarrow s100A \rightarrow s1000A \rightarrow s10000A \rightarrow s100000A \rightarrow s100000e$

40

Die als Wörter der Sprache L erzeugten Binärmuster können als beliebige jedoch eindeutige Darstellung von Datentypen, z.B. als Buchstaben, Zahlen, alphanumerische Zeichen oder Zeichenketten gelesen werden. Liest man sie als die Binärdarstellung von Zahlen, so sind die obigen Wörter in Dezimaldarstellung:

- 5 a) \emptyset (leer)
b) 1
c) 2
d) 32.

10

Erläuterung zu II: das NFR-Verfahren zur Herstellung von Monomersequenzen

Das NFR-Verfahren (Niehaus-Feldkamp-Rauhe-Verfahren) ist ein Verfahren zur Herstellung von Monomersequenzen (Oligo- und Polymeren, z.B. Nukleinsäuren), das gewährleistet, daß

15 die mithilfe des Verfahrens hergestellten Monomersequenzen:

- a) eine eindeutige und einzigartige Abfolge von Monomeren haben;
b) zueinander möglichst unähnlich sind und daher möglichst keine Fehlhybridisierungen untereinander eingehen;
c) bestimmte strukturelle Eigenschaften aufweisen, z.B. ein bestimmtes Verhältnis der
20 verschiedenen Monomere zueinander, das Enthaltensein oder Nicht-Enthaltensein bestimmter Teilsequenzen und eine maximale Übereinstimmung (*Homologie*) untereinander;
d) bestimmte chemische Eigenschaften aufweisen, z.B. das Vorkommen bestimmter chemisch aktiver Teilsequenzen, die Einfluß auf chemische Reaktionen, im Falle von
25 Nukleinsäuren insbesondere die Wechselwirkung mit Proteinen haben (Proteinbindungsstellen, Restriktionsschnittstellen, Stopcodons);
e) bestimmte physikalische Eigenschaften aufweisen, z.B. eine bestimmte Schmelztemperatur.

- 30 Das NFR-Verfahren erlaubt die Herstellung eindeutiger, möglichst unähnlicher Monomersequenzen mit vordefinierbaren strukturellen, chemischen und physikalischen Eigenschaften. Es eignet sich zur Herstellung von Monomersequenzen für kontrollierte chemische Reaktionen, wie sie z.B. für eine molekulare Informationsverarbeitung benötigt werden. Es wird in Schritt III des erfindungsgemäßen Verfahrens für die Implementierung
35 regulärer Grammatiken verwendet.

- Das NFR-Verfahren ist eine erfindungsgemäße Fortbildung des Niehaus-Verfahrens zur Erzeugung eindeutiger, möglichst unähnlicher Sequenzen, das in [Jens Niehaus, DNA Computing: Bewertung und Simulation, *Diplomarbeit am Fachbereich Informatik der Universität Dortmund, Lehrstuhl XI*, 116-123, (1998)] beschrieben ist, und worauf hiermit in
40 vollem Umfang Bezug genommen wird.

Das NFR-Verfahren ist insofern eine erfindungsgemäße Fortbildung des Niehaus-Verfahrens, als erst das NFR-Verfahren die Herstellung von Monomersequenzen erlaubt, wie sie für eine Herstellung in vitro, z.B. zur Implementierung von Grammatiken, benötigt werden. Die Gründe sind im Einzelnen:

- 5 a) Das Niehaus-Verfahren beschreibt nur die Konstruktion von Sequenzen aus Basissequenzen der Länge 6. Da die Länge von Basissequenzen die maximal zwischen verschiedenen Monomersequenzen möglichen Überlappungen beschreibt, daher unmittelbaren Einfluß auf das Hybridisierungsverhalten von Sequenzen und das Funktionieren der erfindungsgemäßen Schritte IV und V hat, muß das Verfahren jedoch für
10 Basissequenzen beliebiger Länge verallgemeinert werden, was im NFR-Verfahren vorgenommen wurde.
- b) Das Niehaus-Verfahren erlaubt nur die Erzeugung von Sequenzen gleicher Länge, wohingegen das NFR-Verfahren Sequenzen unterschiedlicher Länge und unterschiedlicher Anzahl erzeugen kann. Letzteres ist für ein korrektes Hybridisierungsverhalten von
15 Sequenzen und das Funktionieren der erfindungsgemäßen Schritte IV und V unerlässlich.
- c) Das Niehaus-Verfahren gewährleistet Eindeutigkeit und maximale Unähnlichkeit von Sequenzen nur für Sequenzen, die durch eine einmalige Anwendung des Verfahrens erzeugt wurden. Zwingend erforderlich ist jedoch, auch zu schon existierenden Sequenzen jeweils kompatible, d.h. eindeutige und maximal unähnliche Sequenzen erzeugen zu
20 können. Das NFR-Verfahren kann im Gegensatz zum Niehaus-Verfahren zu beliebig vorgegebenen Sequenzen kompatible (d.h. eindeutige und maximal unähnliche) neue Sequenzen erzeugen. Dabei kann das Verfahren beliebig oft auf dieselbe Menge von Sequenzen angewendet werden.
- d) Das Niehaus-Verfahren beschränkt zwar die maximale Länge gemeinsamer
25 Teilsequenzen, jedoch nicht deren Anzahl. Daher können 2 beliebige nach dem Niehaus-Verfahren konstruierte Sequenzen mehrere gemeinsame Teilsequenzen beinhalten, was ungewollt zu einer hohen Homologie (Sequenzübereinstimmung) führen kann. Die Homologie hat ihrerseits unmittelbaren Einfluß auf das Hybridisierungsverhalten von Sequenzen und das Funktionieren der erfindungsgemäßen Schritte IV und V. Das NFR-
30 Verfahren dagegen führt bei der Konstruktion von Sequenzen auch einen Homologievergleich durch und vermeidet dadurch ungewollte Fehlhybridisierungen.
- e) Das Niehaus-Verfahren erlaubt nicht die Integration bestimmter, vorgegebener Sequenzen und Teilsequenzen. Solche Sequenzen sind jedoch für die Ausführung und Kontrolle bestimmter chemischer Reaktionen, z.B. kontrollierter enzymatischer Wechselwirkungen,
35 zwingend notwendig. Beispiele für solche Sequenzen sind im Falle von Nukleinsäuren Proteinbindungsstellen, Restriktionsschnittstellen und Stopcodons. Dagegen erlaubt das NFR-Verfahren die Integration beliebiger Sequenzen und Teilsequenzen in die Herstellung von Monomersequenzen und gewährleistet gleichzeitig Eindeutigkeit und maximale Unähnlichkeit.
- 40 f) Das Niehaus-Verfahren sieht keine Möglichkeit vor, Sequenzen mit bestimmten strukturellen, chemischen oder physikalischen Eigenschaften zu erzeugen. Dagegen

enthält das NFR-Verfahren die Möglichkeit, Sequenzen mit bestimmten strukturellen, chemischen und physikalischen Eigenschaften herzustellen. Dazu zählen u.a. das Verhältnis verschiedener Monomere zueinander und die Schmelztemperatur. Auch dies ist für die Implementierung von Grammatiken in vitro eine unbedingte Voraussetzung.

- 5 g) Das Niehaus-Verfahren erlaubt keine direkte Erzeugung regelrepräsentierender Sequenzen, wie sie zur Implementierung von Grammatiken benötigt werden. Dagegen ermöglicht das NFR-Verfahren die Erzeugung symbolrepräsentierender Sequenzen, die zu den für die Implementierung einer Grammatik benötigten regelrepräsentierenden Sequenzen verknüpft werden können.
- 10 h) Nur mithilfe des NFR-Verfahrens können Sequenzen auch so konstruiert werden, daß die Eigenschaften der Eindeutigkeit, maximalen Unähnlichkeit sowie strukturelle, chemische und physikalische Eigenschaften auch für Teilsequenzen gelten. Z.B. lassen sich im Falle von Nukleinsäuren Monomersequenzen so herstellen, daß sowohl die Gesamtsequenz, als auch z.B. das erste Drittel der Sequenz einen 50%igen GC-Anteil haben. Diese
- 15 Eigenschaft ist z.B. für das Funktionieren der Verkettung von Sequenzen, die pro Molekül für mehrere Hybridisierungsereignisse vorgesehen sind (wie z.B. die im Folgenden beschriebenen Algomere), eine unbedingte Voraussetzung. Nur so lassen sich z.B. Sequenzen so konstruieren, daß unterschiedliche Hybridisierungsereignisse pro Molekül gleichwahrscheinlich sind.
- 20 i) Nur das NFR-Verfahren erlaubt die unmittelbare, automatische Herstellung von Monomersequenzen, z.B. durch einen Oligonukleotidsynthesizer.
- j) Bei der Verknüpfung von Sequenzen zu länger-kettigen Sequenzen kann es zu Verletzungen der Uniqueness kommen. Für die korrekte Herstellung von Polymeren durch das Verknüpfen von Oligomeren (wie z.B. das Verknüpfen von Algomeren zu Logomeren) ist es unbedingt erforderlich, daß diese Uniqueness-Verletzungen nur in einem
- 25 vordefinierten Bereich auftreten, damit keine unkontrollierten Fehlhybridisierungen auftreten (siehe Abbildung 21). Dieses Problem tritt ganz grundsätzlich auf und muß insbesondere für die Verknüpfung von Algomeren und die Verknüpfung von Variablen und Terminalen gelöst werden, weil sonst eine Übersetzung von Grammatiken in Moleküle unmöglich ist. Das Niehaus-Verfahren löst dieses Problem nicht und ist daher für die
- 30 Übersetzung von Grammatiken in Moleküle ungeeignet. Das Problem wird unter Verwendung des NFR-Verfahrens durch eine "Parallel-Extension" Strategie gelöst. Diese wird im Folgenden erläutert.

35 Zur Herstellung von Monomersequenzen werden diese durch das NFR-Verfahren konstruiert und synthetisiert. Zur Herstellung von n Monomersequenzen wird das NFR-Verfahren wie im Folgenden beschrieben durchgeführt, wobei zur Beschreibung des Verfahrens folgende Abkürzungen benutzt werden:

40

- A := Alphabet der Mächtigkeit $a \in \mathbb{N}$;
Im Falle von DNA ist $A := \{a, c, g, t\}$ und $a = 4$.
- 5 S_i := Sequenz.
Folge von Elementen des Alphabets A, das einer Sequenz aus Monomeren entspricht.
- l_{seq} := Länge der pro Verfahrenszyklus zu konstruierenden Sequenzen.
- S_{seq} := Sequenz der Länge l_{seq} .
- $S_{seq,k}$:= (k-1)-tes Monomer der Sequenz S_{seq} ($k \in \mathbb{N}$; $k \leq l_{seq}$).
- 10 l_{bas} := Länge der pro Verfahrenszyklus zur Konstruktion von Sequenzen verwendeten Basissequenzen.
Es gilt: $0 < l_{bas} \leq l_{seq}$.
- S_{bas} := Sequenz der Länge l_{bas} = Basissequenz; Sequenzen der Länge l_{seq} werden aus Basissequenzen konstruiert.
- 15 l_{ov} := maximale Länge der Kette aufeinanderfolgender Monomere, die jeweils zwei durch das Verfahren erzeugte Sequenzen gemeinsam haben („overlap“).
= $l_{bas} - 1$.
- l_{ov}/l_{seq} = Verhältnis von maximal erlaubter Sequenzwiederholung zu Sequenzgesamtlänge;
20 1. Maß für die Fehlhybridisierungswahrscheinlichkeit.
- l_{bas}/l_{seq} = Verhältnis von Basissequenzlänge zu Sequenzgesamtlänge;
2. Maß für die Fehlhybridisierungswahrscheinlichkeit.
- M_{seq} := Menge von Sequenzen.
- M_{bas} := Menge aller Basissequenzen der Länge l_{bas} .
- 25 M_{nobas} := Teilmenge der Menge von Basissequenzen der Länge l_{bas} , die durch Dekomposition aus M_{seq} erhalten wird.
- $|M_{nobas}|$:= Mächtigkeit von M_{nobas} = Anzahl der Basissequenzen in M_{nobas} .
- n := Anzahl pro Verfahrenszyklus zu konstruierender Sequenzen.
- n_{tot} := Anzahl insgesamt in allen Verfahrenszyklen hergestellter Sequenzen.
- 30 n_{max} := maximale Anzahl pro Verfahrenszyklus konstruierbarer Sequenzen.
- h := Homologie. Ein Maß für die Übereinstimmung zwischen zwei Sequenzen.
Es gilt: $0 \leq h \leq 1$.
- t_m := Schmelztemperatur. Für eine DNA-Sequenz wird die Schmelztemperatur nach dem *Nearest Neighbour* Verfahren (siehe Breslauer, K.J., Frank, R., Blocker, H., Marky, L.A., Proc. Natl. Acad. Sci., **83**, 3746-3750, 1989) bestimmt:
35 $t_m = \Delta H / (\Delta S + R \times \ln(C/4)) - 273,15^\circ\text{C}$
Dabei sind ΔH und ΔS Enthalpie und Entropie der DNA-Helix, R die molare Gaskonstante und C die Konzentration der DNA-Sequenz.
- 40

Das Herstellungsverfahren erfolgt als ein Abfolge beliebig, jedoch endlich vieler Verfahrenszyklen, in denen n_{tot} Sequenzen konstruiert werden, und einer nachfolgenden Synthese, in der alle n_{tot} Sequenzen in vitro erzeugt werden.

Die maximale Anzahl n_{max} der pro Verfahrenszyklus konstruierbaren Sequenzen errechnet sich entsprechend den oben gegebenen Definitionen nach

$$n_{\text{max}} = f(a, l_{\text{bas}}, l_{\text{seq}}) := \lfloor \left(\frac{1}{2} * ((a^{l_{\text{bas}}}) - (a^{l_{\text{bas}}/2}) - |M_{\text{nobas}}|) \right) / (l_{\text{seq}} - l_{\text{ov}}) \rfloor.$$

D.h. es lassen sich maximal n_{max} Sequenzen der Länge l_{seq} aus Elementen eines Alphabetes A der Größe a (im Falle von DNA ist $a = 4$) konstruieren, die in maximal l_{ov} zusammenhängenden Ketten von Monomeren übereinstimmen. Die Anzahl der tatsächlich pro Verfahrenszyklus erhaltenen Sequenzen kann geringer sein, wenn diese bestimmte physikalische, chemische oder strukturelle Eigenschaften aufweisen sollen, oder wenn zusätzlich zu bereits vorhandenen Sequenzen weitere Sequenzen hergestellt werden.

Im Einzelnen sind folgende Verfahrensschritte nötig:

15

- 1 Es werden n_{tot} Sequenzen in z Verfahrenszyklen konstruiert und in einer Menge M_{seq} gesammelt. Man beginnt mit der leeren Sequenzmenge M_{seq} .
- 2 Zur Sequenzmenge M_{seq} können einmalig, vor Beginn des Verfahrenszyklus, beliebige, jedoch eindeutige, Sequenzen hinzugefügt werden. Dies kann nützlich sein, um
20 bestimmte Sequenzen hinzuzufügen, die aus Gründen der weiteren Anwendung in der Menge der hergestellten Monomersequenzen enthalten sein sollen. Es empfiehlt sich dabei, nicht zu viele und nicht zu lange Sequenzen hinzuzufügen, da das Verfahren für diese Sequenzen nicht dieselben Eigenschaften garantiert wie für die im Folgenden konstruierten Sequenzen.
- 25 3 Beginn des Verfahrenszyklus: Es werden die strukturellen, chemischen und physikalischen Eigenschaften festgelegt, die die herzustellenden Sequenzen erfüllen müssen. Strukturelle Eigenschaften sind zumindest das Enthaltensein oder Nicht-Enthaltensein bestimmter Teilsequenzen (ggfs. auch die Positionen, an denen die Teilsequenzen in den zu erzeugenden Sequenzen enthalten oder nicht enthalten sein
30 sollen) und das Verhältnis der Anzahl der verschiedenen Monomere zueinander (im Falle von DNA: Verhältnis GC/AT). Chemische Eigenschaften sind mindestens das Vorkommen bestimmter Teilsequenzen, die Einfluß auf die chemische Wechselwirkung mit eigenen oder anderen Substanzen haben (im Falle von DNA z.B. bestimmte Proteinbindungsstellen, Restriktionsschnittstellen wie z.B. 5'gatatc3' für EcoRV, die
35 Stopcodons 5'tca3', 5'tta3', 5'cta3' usw.). Zu den festzulegenden physikalischen Eigenschaften zählt zumindest die Schmelztemperatur t_m der herzustellenden Sequenzen.

- 4 Es werden jetzt die Anzahl der pro Zyklus gewünschten Sequenzen $n \leq n_{\max}$, nach obiger Formel l_{seq} der zu konstruierenden Sequenzen und l_{bas} der zur Konstruktion benötigten Basissequenzen festgelegt. l_{seq} und l_{bas} werden nach der oben angegebenen Formel so gewählt, daß die gewünschte Anzahl n zu konstruierender Sequenzen erreicht werden kann. Da $l_{\text{bas}}/l_{\text{seq}}$ proportional zur Fehlhybridisierungswahrscheinlichkeit ist, werden l_{bas} und l_{seq} so gewählt, daß das Verhältnis $l_{\text{bas}}/l_{\text{seq}}$ möglichst gering ist (Werte unter 0,3 sind empfehlenswert) und Werte für l_{seq} gute Hybridisierungsbedingungen garantieren. Für DNA sind temperaturabhängig beliebige Werte für $l_{\text{seq}} > 0$ möglich.
- 5 Es wird die Menge M_{bas} aller Sequenzen des Alphabetes A der Länge l_{bas} erzeugt. Die solcherart erzeugten Sequenzen werden als *Basissequenzen* bezeichnet. Es gibt dabei immer $a^{l_{\text{bas}}}$ Basissequenzen. Diese Basissequenzen dienen der Konstruktion der herzustellenden Sequenzen. Jede Basissequenz bekommt einen Status „benutzt“ oder „unbenutzt“ zugewiesen. Es werden zunächst alle als unbenutzt markiert.
- 6 Selbst-komplementäre Basissequenzen aus M_{bas} werden nun als benutzt markiert. Dabei gibt es genau $a^{l_{\text{bas}}/2}$ zu sich selbst komplementäre Basissequenzen.
- 7 Sollten bereits Sequenzen in der Menge M_{seq} vorhanden sein, so können entweder dazu kompatible neue Sequenzen konstruiert werden oder die Sequenzen einer Teilmenge von M_{seq} kompatibel verlängert werden. Dazu wird innerhalb eines Dekompositionsverfahrens aus den bereits vorhandenen Sequenzen aus M_{seq} eine Menge M_{nobas} aller Teilsequenzen der in Schritt 4 vorgegebenen Länge l_{bas} der Sequenzen aus M_{seq} gebildet:
- Setze $M_{\text{nobas}} = \{\}$.
- Jede Sequenz S_{seq} aus M_{seq} wird nun in $(l_{\text{seq}} - l_{\text{ov}})$ Dekompositionsschritten zerlegt:
- Beginne mit $i = 0$ mit dem Dekompositionsschritt:
- 25 Solange $i < (l_{\text{seq}} - l_{\text{ov}})$:
- Bilde als neue Sequenz S_{neu} als Sequenz der Länge l_{bas} bestehend aus den Monomeren $S_{\text{seq},i}$ bis $S_{\text{seq},i+l_{\text{bas}}}$: $S_{\text{neu}} := S_{\text{seq},i}, \dots, S_{\text{seq},i+l_{\text{bas}}}$.
- Füge S_{neu} der Menge M_{nobas} hinzu.
- Setze $i := i + 1$.
- 30 Die auf diese Weise erhaltene Menge M_{nobas} ist definitionsgemäß eine Teilmenge von M_{bas} .
- 8 Alle Basissequenzen der Menge M_{bas} , die auch in M_{nobas} vorkommen, werden als benutzt markiert, so daß für die Konstruktion von Sequenzen genau $(a^{l_{\text{bas}}} - a^{l_{\text{bas}}/2} - |M_{\text{nobas}}|)$ unbenutzte Basissequenzen übrig bleiben, aus denen nun maximal

$$n_{\max} = \left\lfloor \frac{\left(\frac{(a^{l_{\text{bas}}}) - (a^{l_{\text{bas}}/2}) - |M_{\text{nobas}}|}{2} \right)}{(l_{\text{seq}} - l_{\text{ov}})} \right\rfloor$$

verschiedene neue Sequenzen konstruiert werden können, die untereinander und zu den bereits vorhandenen Sequenzen keine Basissequenzen bzw. deren Komplemente gemeinsam haben.

- 9 Aus den Basissequenzen wird ein gerichteter Graph konstruiert, dessen Knoten
5 bestimmte Basissequenzen repräsentieren: Der Graph enthält genau $a^{l_{bas}}$ Knoten, von denen bereits $(a^{l_{bas/2}} + |M_{nobas}|)$ Knoten als benutzt markiert sind. Jeder Knoten ist mit einer Basissequenz $S_{b(i)}$ assoziiert, die nicht zu sich selbst komplementär ist (im weiteren wird es keine Unterscheidung von Knoten und assoziierter Basissequenz geben).

Es existiert eine Kante von $S_{b(i)}$ nach $S_{b(k)}$, wenn die l_{ov} letzten Buchstaben von $S_{b(i)}$ den l_{ov} ersten Buchstaben von $S_{b(k)}$ entsprechen, wenn also gilt:

$$S_{b(i),2}, \dots, S_{b(i),l_{bas}} = S_{b(k),1}, \dots, S_{b(k),l_{bas}-1}.$$

Der Knoten $S_{b(k)}$ wird *Nachfolgeknoten* von $S_{b(i)}$ genannt.

Der Knoten, der mit dem Komplement der Basissequenz assoziiert ist, die durch den Knoten $S_{b(i)}$ kodiert wird, wird *Komplementknoten* zu Knoten $S_{b(i)}$ genannt.

- 15 Sequenzen der Länge l_{seq} werden gefunden, indem ein Pfad mit $(l_{seq} - l_{ov})$ Knoten gesucht wird. Jeder Knoten darf maximal einmal benutzt werden. Außerdem darf für jeden Knoten, der in einem der Pfade liegt, der Komplementknoten in keinem Pfad vorkommen. Der Anfangsknoten des Pfades trägt wie die Sequenz einen Status „benutzt“ bzw. „unbenutzt“.

- 20 10 Aus dem Graphen konstruiert man Sequenzen nach folgendem Verfahren:

Kennzeichne alle unbenutzten Knoten des Graphen als unbenutzte Anfangsknoten.

Für jedes s zwischen 1 und n :

Solange Knoten $S_{b(k)}$ existiert, der noch nicht als benutzter Anfangsknoten markiert ist:

- 25 Wähle einen unbenutzten Knoten $S_{b(k)}$ aus.

Markiere Knoten $S_{b(k)}$ als benutzten Anfangsknoten; außerdem:

Markiere Sequenz $S_{b(k)}$ und sein Komplement als benutzt.

Nun wird eine neue Sequenz S_{neu} konstruiert, die neues Element für M_{seq} ist oder eine bestehenden Sequenz S_{sub} aus M_{seq} verlängert:

- 30 Setze $S_{neu,0} := S_{b(k),1}$.

Falls sie als neue Sequenz konstruiert wird, setze $i := 0$.

Falls sie als Verlängerung einer bereits bestehenden Sequenz S_{sub} konstruiert wird, setze $i := l_{sub}$.

Solange $i < l_{seq} - (l_{ov}-1)$ und $i \geq 0$ gilt:

- 35 Existiert kein unbenutzter Nachfolgeknoten $S_{b(m)}$ von Knoten $S_{b(k)}$, so markiere Knoten $S_{b(k)}$ und dessen Komplementknoten als unbenutzt und setze $i := i-1$.

Sonst wähle per Zufall einen unbenutzten Nachfolgeknoten $S_{b(m)}$ von Knoten $S_{b(k)}$ aus und markiere b_m und dessen Komplementknoten als benutzt. Setze zusätzlich: $i := i+1$, $k := m$, $S_{neu,i} := S_{b(m),1}$.

40

21

Wenn $i = l_{seq} - (l_{ov}-1)$ ist, ist die Sequenz S_{neu} fertig: sie besteht aus den Buchstaben $S_{neu,0}$ bis $S_{neu,(l_{seq}-l_{ov})}$.

- 11 Zu jeder der in Schritt 10 erhaltenen Sequenzen werden die jeweiligen strukturellen, chemischen und physikalischen Eigenschaften im Vergleich zu
- 5 den in Schritt 3 festgelegten Bedingungen ermittelt. Genügt die jeweilige Sequenz nicht den vorgegebenen Bedingungen, so wird sie gelöscht und die von ihr benutzten Knoten und Komplementknoten wieder als unbenutzt markiert.
- 12 Wenn die Anzahl der konstruierten Sequenzen nicht ausreicht, kann der
- 10 Verfahrenszyklus von Schritt 3 oder 4 an beliebig oft wiederholt werden. Insbesondere ist es möglich, die strukturellen, chemischen und physikalischen Kriterien und die Werte für n , l_{seq} , l_{bas} pro Verfahrenszyklus jeweils neu zu setzen. So ist es möglich, Sequenzen mit unterschiedlichen strukturellen, chemischen und physikalischen Eigenschaften, sowie unterschiedlicher Länge zu konstruieren, die trotzdem kompatibel, also eindeutig und
- 15 maximal unähnlich sind.
- Ansonsten geht man zum nächsten Schritt, der in vitro Synthese.
- 13 Die konstruierten Sequenzen werden in vitro, im Falle von Nukleinsäuren vorzugsweise mittels eines Oligonukleotidsynthesizers (z.B. ABI 392, ABI 398, ABI 3948 von Perkin-Elmer Applied Biosystems), synthetisiert. Dazu werden die Sequenzdaten der
- 20 Steuerungseinheit des Oligonukleotidsynthesizers übermittelt und die Sequenzen als einzelsträngige Nukleinsäuren hergestellt. Die Sequenzen können auch kommerziell bestellt werden. Zweckmäßig sind PAGE-gereinigte Oligonukleotide (z.B. von ARK Scientific GmbH Biosystems, 64293 Darmstadt, erhältlich).
- 25 Um Grammatiken in Moleküle übersetzen zu können muß das oben erwähnte Problem der Uniqueness-Verletzung bei der Verknüpfung von Sequenzen gelöst werden (siehe Abbildung 21). Dieses Problem ist mit dem Niehaus-Verfahren nicht zu behandeln, da dort keine Verknüpfung von Sequenzen vorgesehen ist. Im Einzelnen ist das Niehaus-Verfahren für die Lösung folgender Teilprobleme nicht mächtig genug:
- 30 - Die Verknüpfung von Sequenzen kann zu einer Uniqueness-Verletzung führen (siehe Abbildung 21).
- Bei der Verknüpfung von Variablen und Terminalen kann die Sequenz einer Variable je nach Grammatik in mehr als eine Terminalsequenz pro Ende übergehen und umgekehrt (siehe Abbildung 22 und Abbildung 24).
- 35 - Bei der Verknüpfung z.B. mehrerer Terminalen mit der selben Variable kann es zu Uniqueness-Verletzungen kommen, die sich durch einfache Pfadsuche nicht auflösen lassen (siehe Abbildung 23).

Das Problem wird mit einer als "Parallel-Extension" bezeichneten Strategie unter Verwendung

40 des NFR-Verfahrens gelöst. Dabei wird im Einzelnen folgendermaßen vorgegangen

(beispielhaft gezeigt für die Konstruktion von Variablen zu vorgegebenen Terminalen, siehe auch Abbildung 22 bis Abbildung 24):

1. Eine Gruppe von Sequenzen (hier: die Terminalen, z.B. Sequenzen zur Repräsentation von Bits oder Sequenzen mit bestimmten chemischen oder biologischen Eigenschaften) seien vorgegeben.
2. Sammle für jede Variable alle Paare von Terminalen, deren Sequenzen die jeweilige Variablensequenz im Logomer einrahmen werden und fasse sie zu Gruppen zusammen, eine Gruppe für jede Variable (siehe Abbildung 22 und Abbildung 24). Ordne die Pfade der Terminalpaare so an, daß zwischen jedem Paar eine Lücke bleibt, deren Länge der Variablenpfadlänge ($l_{seq} - l_{bas} + 1$), plus jeweils $l_{bas} - 1$ Knoten für jeden der beiden Übergänge entspricht. Dabei ist ein Übergang derjenige Pfad der Länge $l_{bas} - 1$, der jeweils zusammengehörende Terminalen und Variablen miteinander verbindet. In Abbildung 21 besteht dieser z.B. aus den markierten Basissequenzen. Übergänge können zusammenlaufen (z.B. die Terminalen a, b, c zur Variable A in Abbildung 22) oder sich verzweigen (z.B. die Variable A zu den Terminalen b, c, d in Abbildung 22).
3. Wähle die jeweils letzte Basissequenz jeder Terminalsequenz auf der linken Seite als Startknoten für den jeweiligen Pfad (siehe Abbildung 24).
4. Erzeuge die Terminal-Variable-Übergänge simultan, d.h. suche in jedem Iterationsschritt für alle Pfade je einen Nachfolgeknoten, dessen letztes Monomer für alle Pfade einer Gruppe das selbe ist. Kann in diesem Schritt kein unbenutzter und erlaubter Knoten gefunden werden, so wird Backtracking ausgelöst. Ist dieses nicht möglich, so scheitert die Übersetzung einer Grammatik in Algomere an diesem Punkt und die Übersetzung muß mit veränderten (vorzugsweise weniger restriktiven) Parametern wiederholt werden. Zulässig ist dagegen die mehrfache Verwendung eines Knotens in einem Iterationsschritt für die jeweils zusammengehörenden Übergänge (siehe Abbildung 23 und Abbildung 24).
5. Erzeuge die Variablen-Sequenzen ebenfalls simultan als Verlängerung der Übergänge.
6. Erzeuge die Variable-Terminal-Übergänge analog wie unter 4 beschrieben. Hierbei sind die Nachfolger jedoch nicht frei wählbar, sondern durch die $l_{bas} - 1$ ersten Nukleotide der rechten Terminalsequenz vorgegeben. Die Pfadsuche folgt hier dieser Vorgabe um festzustellen, ob die Pfade sich vervollständigen lassen.

Durch das als "Parallel-Extension" bezeichnete Verfahren wird es möglich, eine Schnittstelle zwischen Computer und den in-vitro ausgeführten Schritten, beginnend mit der Synthetisierung von Oligonukleotiden, zu schaffen und die Schritte von Definition von Grammatiken bis zur Herstellung von Molekülen in-vitro vollständig zu automatisieren.

Erläuterung zu III: Implementierung regulärer Grammatiken mit dem NFR-Verfahren

Die Herstellung von Monomersequenzen zur Darstellung der Regelmenge von Grammatiken in Schritt III des erfindungsgemäßen Verfahrens erfolgt mit Hilfe des NFR-Verfahrens (Niehaus-Feldkamp-Rauhe-Verfahren) nach Schritt II des erfindungsgemäßen Verfahrens. Zur Implementierung von Grammatiken werden nach dem NFR-Verfahren für r Regeln einer Grammatik G genau $2r$ Monomersequenzen hergestellt, so daß die Monomersequenzen für die s dargestellten Symbole (Terminale und Variablen) und die Regeln R der Grammatik G eindeutig und zueinander möglichst unähnlich sind, sowie die geforderten strukturellen, chemischen und physikalischen Eigenschaften aufweisen.

Die Monomersequenzen werden dazu nach dem NFR-Verfahren so hergestellt, daß sie wie im Folgenden beschrieben zu Algomeren zusammengesetzt werden können. Es werden dabei jeweils 2 Monomersequenzen zu einem Algomer zusammengesetzt, das genau eine Regel R einer Grammatik G repräsentiert. Dabei enthalten beide Monomersequenzen jeweils Sequenzen, die die nach der Regel R erforderlichen zusammengehörenden Symbole (Terminale oder Variablen) enthalten.

Der Entwurf der Oligomere nach Schritt III des erfindungsgemäßen Verfahrens ist abhängig von der chemischen Natur der eingesetzten Monomere. Im bevorzugten Fall der aus Nukleotiden aufgebauten Oligomere wird folgendermaßen vorgegangen:

Algomere sind doppelsträngig und haben einen 5'-Strang (*Oberstrang*) und einen 3'-Strang (*Unterstrang*). Oberstrang und Unterstrang können die Verknüpfung jeweils einer Terminal- und einer Variablensequenz sein.

Es seien X , Z beliebige Variablen, S eine Variable, die als Startsymbol fungiert, y , z beliebige Terminalsymbole. Dann wird für jede Regel der Form:

a) $S := yX$

ein Algomer konstruiert, das eine eindeutige doppelsträngige Kernsequenz y enthält, an die am 3'-Ende eine eindeutige einzelsträngige Überhangsequenz X angehängt ist; " $:=$ " bedeutet in diesem Fall, daß S und yX identisch sind, d.h. yX als Startermolekül fungiert.

b) $X \rightarrow yZ$

ein Algomer konstruiert, das eine eindeutige doppelsträngige Kernsequenz y enthält, an die am 5'-Ende eine eindeutige einzelsträngige Überhangsequenz X und am 3'-Ende eine eindeutige einzelsträngige Überhangsequenz Z angehängt ist. X kann dabei identisch mit Z sein.

c) $X \rightarrow z$

ein Algomer konstruiert, das eine eindeutige doppelsträngige Kernsequenz z enthält, an die am 5'-Ende eine eindeutige einzelsträngige Überhangsequenz X angehängt ist.

Algomere der Form $S := yX$ und $X \rightarrow z$ heißen *Terminatoren* ("Start", "Ende"), da sie zum Kettenabbruch während der Polymerisation im erfindungsgemäßen Verknüpfungsschritt V führen, der im Folgenden beschrieben wird; Algomere der Form $X \rightarrow yZ$ heißen *Elongatoren*,

weil sie zu einer Kettenverlängerung während der Polymerisation im erfindungsgemäßen Verknüpfungsschritt V führen.

- Vorzugsweise umfassen die in den Schritten II und III des erfindungsgemäßen Verfahrens hergestellten Monomersequenzen Nukleotide, insbesondere Ribonukleotide, besonders bevorzugt Desoxyribonukleotide. Die in Schritt II des erfindungsgemäßen Verfahrens konstruierten Monomersequenzen können vorteilhafterweise bestimmte Sequenzen, z.B. Stopcodons, Erkennungssequenzen für Nukleinsäuren spaltende Enzyme (Restriktionsnukleasen), Erkennungssequenzen für Nukleinsäure-bindende Proteine u.a. enthalten, wie sie beispielsweise in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)], [Rolf Knippers, Molekulare Genetik, Georg Thieme Verlag, (1997)] und [Benjamin Lewin, Genes V, Oxford University Press, (1994)] beschrieben werden.

15 Erläuterung zu IV: Algomer-Assemblierung

- Das Assemblieren der in Schritt III synthetisierten Sequenzen erfolgt in Schritt IV des erfindungsgemäßen Verfahrens. Im Rahmen der vorliegenden Erfindung sind zur Durchführung des Schrittes IV alle dem Fachmann bekannten Techniken, Oligomersequenzen zusammenzusetzen, einsetzbar. Im erfindungsgemäß bevorzugten Fall des Einsatzes von Nukleotidsequenzen ist es zweckmäßig folgendermaßen vorzugehen:

- a) Die zu Elongatoren gehörenden einzelsträngigen Sequenzen werden phosphoryliert. Dazu sind verschiedene dem Fachmann bekannte Protokolle möglich, z.B. das Folgende:
- 25 In einem 20 μ l Ansatz werden 16 μ l einer synthetisierten 100 μ M Sequenz, 2 μ l Ligationspuffer (z. B. 50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 10 mM Dithiothreitol, 1 mM ATP, 25 μ g/ml BSA Bovine Serum Albumin, New England Biolabs) und 2 μ l PNK (Polynukleotidkinase, z.B. von New England Biolabs, Katalognr. #201S oder #201L) 1 Stunde bei 37°C inkubiert. Volumina, Inkubationszeit und -temperatur können in dem Fachmann bekannter Weise variiert werden.
- 30 b) Jeweils in einem Ansatz werden die zu einem Algomer gehörenden Einzelstränge (Ober- und Unterstrang) zu einem fertigen Algomer hybridisiert. Für Elongatoren können einfach die Phosphorylierungsansätze von Ober- und Unterstrang verwendet werden. Für die Hybridisierung sind mehrere Protokolle möglich; bevorzugt ist ein Denaturierungsschritt bei etwa 95°C zu Anfang (dieser deaktiviert gleichzeitig die PNK in den Ansätzen der Elongatoren) und eine langsame Hybridisierung. Z.B. kann für Oligomere der Länge 30 folgendes Protokoll verwendet werden, das gemäß den Kenntnissen des Fachmanns variiert werden kann:
- 35 In einem 40 μ l Ansatz werden jeweils 20 μ l des Oberstranges (100 μ M) und 20 μ l des Unterstranges (100 μ M) in einem Thermozykler 5 Minuten auf 95°C erhitzt, 5 Minuten auf 72°C
- 40 inkubiert und dann 25 Minuten jeweils 1°C pro Minute abgekühlt.

Erläuterung zu V: Herstellung von Logomeren durch Verkettung von Algomeren: Symbolpolymerisation

In Schritt V des erfindungsgemäßen Verfahrens werden die Algomere zu längerkettigen, informationstragenden Polymeren (Logomeren) verknüpft. Da Algomere die Regelmenge R einer Grammatik G darstellen, entsprechen alle dabei erzeugten Logomere Wörtern der Sprache $L(G)$, die durch die Grammatik G beschrieben wird.

Der geregelte Prozeß der Verkettung von Algomeren zu Logomeren wird als „Symbolpolymerisation“ bezeichnet, im Falle, daß die Terminalsymbole Bits repräsentieren, das heißt, im Falle, daß gilt:

$$\Sigma := \{0, 1, s_0, s_1, s_2, \dots, s_{n-1}, s_n, e_0, e_1, e_2, \dots, e_{m-1}, e_m\} \text{ mit } n, m \in \mathbb{N} \text{ und } n, m \geq 0,$$

wird der Prozeß als „Bitpolymerisation“ bezeichnet.

Im Rahmen der vorliegenden Erfindung sind zur Durchführung des Schrittes V alle dem Fachmann bekannten Techniken, Oligomere zu Polymeren zu verknüpfen, einsetzbar. Im erfindungsgemäß bevorzugten Fall des Einsatzes von Oligonukleotiden ist es zweckmäßig, die zu verknüpfenden Algomere in Gegenwart von Ligase zu inkubieren. Beispielsweise kann nach folgendem Protokoll vorgegangen werden, das gemäß den Kenntnissen des Fachmanns variiert werden kann:

In einem 27 μ l Ansatz werden 1 μ l 50 μ M „Start“-Algomer, 1 μ l 50 μ M „Ende“-Algomer und 2x jeweils 10 μ l 40 μ M Elongator-Algomere (aus Schritt IV des erfindungsgemäßen Verfahrens erhalten), 1,5 μ l 10mM rATP und 3,5 μ l T4 DNA Ligase mit 400 NEB Units/ μ l (z.B. Katalognr. #202S und #202L NEB) zwischen 4°C und 25°C für 2 bis 24 Stunden inkubiert.

Andere Reaktionsvolumina funktionieren analog. Inkubationszeit und -temperatur können in dem Fachmann bekannter Weise variiert werden.

Abbildung 3 zeigt das Resultat einer solchen Symbolpolymerisation.

Es können dabei grundsätzlich beliebig viele Algomere, die keine Terminatoren sind, verknüpft werden. Der Polymerisationsvorgang kommt für jedes einzelne Logomer dann zu einem Stillstand, wenn das entsprechende Molekül an seinen Enden jeweils ein Algomer trägt, das als Terminatormolekül („Start“, „Ende“) fungiert.

Vereinzelung und Vervielfältigung von Logomeren durch Klonierung

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein Verfahren zur Isolierung und Vervielfältigung von informationstragenden Polymeren, die nach einem der vorhergehenden Ansprüche erhalten wurden, das dadurch gekennzeichnet ist, daß man in Schritt V erhaltene informationstragende Polymere in Klonierungsvektoren ligiert, kompetente Zellen mit diesen Vektoren transformiert und die erfolgreich transformierten Bakterien anhand von Selektionsmarkern selektioniert.

Erläuterung zu V: Herstellung von Logomeren durch Verkettung von Algomeren: Symbolpolymerisation

In Schritt V des erfindungsgemäßen Verfahrens werden die Algomere zu längerkettigen, informationstragenden Polymeren (Logomeren) verknüpft. Da Algomere die Regelmenge R einer Grammatik G darstellen, entsprechen alle dabei erzeugten Logomere Wörtern der Sprache $L(G)$, die durch die Grammatik G beschrieben wird.

Der geregelte Prozeß der Verkettung von Algomeren zu Logomeren wird als „Symbolpolymerisation“ bezeichnet, im Falle, daß die Terminalsymbole Bits repräsentieren, das heißt, im Falle, daß gilt:

$$\Sigma := \{0, 1, s_0, s_1, s_2, \dots, s_{n-1}, s_n, e_0, e_1, e_2, \dots, e_{m-1}, e_m\} \text{ mit } n, m \in \mathbb{N} \text{ und } n, m \geq 0,$$

wird der Prozeß als „Bitpolymerisation“ bezeichnet.

Im Rahmen der vorliegenden Erfindung sind zur Durchführung des Schrittes V alle dem Fachmann bekannten Techniken, Oligomere zu Polymeren zu verknüpfen, einsetzbar. Im erfindungsgemäß bevorzugten Fall des Einsatzes von Oligonukleotiden ist es zweckmäßig, die zu verknüpfenden Algomere in Gegenwart von Ligase zu inkubieren. Beispielsweise kann nach folgendem Protokoll vorgegangen werden, das gemäß den Kenntnissen des Fachmanns variiert werden kann:

In einem 27 μ l Ansatz werden 1 μ l 50 μ M „Start“-Algomer, 1 μ l 50 μ M „Ende“-Algomer und 2x jeweils 10 μ l 40 μ M Elongator-Algomere (aus Schritt IV des erfindungsgemäßen Verfahrens erhalten), 1,5 μ l 10mM rATP und 3,5 μ l T4 DNA Ligase mit 400 NEB Units/ μ l (z.B. Katalognr. #202S und #202L NEB) zwischen 4°C und 25°C für 2 bis 24 Stunden inkubiert.

Andere Reaktionsvolumina funktionieren analog. Inkubationszeit und -temperatur können in dem Fachmann bekannter Weise variiert werden.

Abbildung 3 zeigt das Resultat einer solchen Symbolpolymerisation.

Es können dabei grundsätzlich beliebig viele Algomere, die keine Terminatoren sind, verknüpft werden. Der Polymerisationsvorgang kommt für jedes einzelne Logomer dann zu einem Stillstand, wenn das entsprechende Molekül an seinen Enden jeweils ein Algomer trägt, das als Terminatormolekül („Start“, „Ende“) fungiert.

Vereinzelung und Vervielfältigung von Logomeren durch Klonierung

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein Verfahren zur Isolierung und Vervielfältigung von informationstragenden Polymeren, die nach einem der vorhergehenden Ansprüche erhalten wurden, das dadurch gekennzeichnet ist, daß man in Schritt V erhaltene informationstragende Polymere in Klonierungsvektoren ligiert, kompetente Zellen mit diesen Vektoren transformiert und die erfolgreich transformierten Bakterien anhand von Selektionsmarkern selektioniert.

- Überstand aufnehmen und 200 μ l Chloroform zugeben, vortexen, bei 10000g 3 Minuten zentrifugieren
- Überstand aufnehmen, mit 1/10 Vol. 3M NaAc ansäuern und mit 2,5x Volumen 100% EtOH versetzen, vortexen, für mindestens 15 Minuten auf -70°C
- 5 - mindestens 15 Minuten auf 10000g zentrifugieren, Überstand abnehmen und verwerfen
- mit ca. 500 μ l 70% EtOH waschen, für 5-10 Minuten auf 10000g zentrifugieren, Überstand abnehmen und verwerfen
- Pellet trocknen und in Aqua dest. wiederaufnehmen, so daß das geschnittene Plasmid mit 100ng/ μ l bis 1 μ g/ μ l vorliegt (höhere Konzentrationen sind auch möglich). Plasmid kann für
10 weitere Ligationen nötigenfalls verdünnt werden.

b) Ligation des Logomers in den Klonierungsvektor

Die in Schritt V des erfindungsgemäßen Verfahrens erhaltenen Logomere und die präparierten Klonierungsvektoren werden ligiert. Durch die Art und Weise der Präparation ist
15 gewährleistet, daß möglichst nur die gewünschten Ligationen stattfinden: Die Terminatoren der Logomere tragen zwei verschiedene Überhangsequenzen, die zu den durch Restriktion entstandenen Überhangsequenzen des Klonierungsvektors kompatibel sind. Dadurch können die Logomere nur in einer definierten Richtung in den Klonierungsvektor ligieren. Außerdem sind die Terminatoren der Logomere nicht phosphoryliert, weshalb Logomere ausschließlich
20 mit den Überhangsequenzen des Klonierungsvektors ligieren können. Die Ligation erfolgt nach einem üblichen Ligationsprotokoll wie z.B. in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)] beschrieben, das gemäß den Kenntnissen des Fachmanns variiert werden kann, beispielsweise wie folgt:

Das molare Verhältnis Logomere/Klonierungsvektor kann z.B. 500 betragen, in 10 μ l Gesamtvolumen werden:

25 1 μ l Plasmid (10ng/ μ l = 5nM)

0,5 μ l 4 μ M Logomere aus Schritt V des erfindungsgemäßen Verfahrens

1 μ l 10x Ligationspuffer

0,5 μ l T4 DNA Ligase (400u/ μ l)

30 7 μ l H₂O

für 12 Stunden auf 16°C ligiert.

Das Protokoll kann gemäß den Kenntnissen des Fachmanns variiert werden. Z.B. ist es möglich, ein Protokoll zur TCL (Temperature Cycle Ligation, [Lund, A.H., Duch, M., Pedersen, F.S, Increased cloning efficiency by temperature-cycle ligation, *Nucleic Acids Research*, 24:
35 (4), 800-801, (1996)]) zu verwenden.

Der erhaltene Ligationsansatz wird nun für die Transformation kompetenter Zellen verwendet, wie nachfolgend beispielhaft gezeigt:

c) Transformation kompetenter Zellen

Die Transformation von Bakterien (kompetente Zellen, Hoststamm z.B. dH5 α , GIBCO BRL, Katalog Nr.: 18258-012) mit dem Ligationsansatz aus b) erfolgt nach einem der üblichen
40 Protokolle, wie z.B. in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A

- Überstand aufnehmen und 200 μ l Chloroform zugeben, vortexen, bei 10000g 3 Minuten zentrifugieren
- Überstand aufnehmen, mit 1/10 Vol. 3M NaAc ansäuern und mit 2,5x Volumen 100% EtOH versetzen, vortexen, für mindestens 15 Minuten auf -70°C
- 5 - mindestens 15 Minuten auf 10000g zentrifugieren, Überstand abnehmen und verwerfen
- mit ca. 500 μ l 70% EtOH waschen, für 5-10 Minuten auf 10000g zentrifugieren, Überstand abnehmen und verwerfen
- Pellet trocknen und in Aqua dest. wiederaufnehmen, so daß das geschnittene Plasmid mit 100ng/ μ l bis 1 μ g/ μ l vorliegt (höhere Konzentrationen sind auch möglich). Plasmid kann für
- 10 weitere Ligationen nötigenfalls verdünnt werden.

b) Ligation des Logomers in den Klonierungsvektor

Die in Schritt V des erfindungsgemäßen Verfahrens erhaltenen Logomere und die präparierten Klonierungsvektoren werden ligiert. Durch die Art und Weise der Präparation ist

15 gewährleistet, daß möglichst nur die gewünschten Ligationen stattfinden: Die Terminatoren der Logomere tragen zwei verschiedene Überhangsequenzen, die zu den durch Restriktion entstandenen Überhangsequenzen des Klonierungsvektors kompatibel sind. Dadurch können die Logomere nur in einer definierten Richtung in den Klonierungsvektor ligieren. Außerdem sind die Terminatoren der Logomere nicht phosphoryliert, weshalb Logomere ausschließlich

20 mit den Überhangsequenzen des Klonierungsvektors ligieren können. Die Ligation erfolgt nach einem üblichen Ligationsprotokoll wie z.B. in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)] beschrieben, das gemäß den Kenntnissen des Fachmanns variiert werden kann, beispielsweise wie folgt:

Das molare Verhältnis Logomere/Klonierungsvektor kann z.B. 500 betragen, in 10 μ l Gesamtvolumen werden:

25 1 μ l Plasmid (10ng/ μ l = 5nM)

0,5 μ l 4 μ M Logomere aus Schritt V des erfindungsgemäßen Verfahrens

1 μ l 10x Ligationspuffer

0,5 μ l T4 DNA Ligase (400u/ μ l)

30 7 μ l H₂O

für 12 Stunden auf 16°C ligiert.

Das Protokoll kann gemäß den Kenntnissen des Fachmanns variiert werden. Z.B. ist es möglich, ein Protokoll zur TCL (Temperature Cycle Ligation, [Lund, A.H., Duch, M., Pedersen, F.S, Increased cloning efficiency by temperature-cycle ligation, *Nucleic Acids Research*, 24:

35 (4), 800-801, (1996)]) zu verwenden.

Der erhaltene Ligationsansatz wird nun für die Transformation kompetenter Zellen verwendet, wie nachfolgend beispielhaft gezeigt:

c) Transformation kompetenter Zellen

Die Transformation von Bakterien (kompetente Zellen, Hoststamm z.B. dH5 α , GIBCO BRL, Katalog Nr.: 18258-012) mit dem Ligationsansatz aus b) erfolgt nach einem der üblichen Protokolle, wie z.B. in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A

40

Laboratory Manual, (1989)] beschrieben wird und das gemäß den Kenntnissen des Fachmanns variiert werden kann, z.B. wie folgt:

- 200µl kompetenter Zellen ($\sim 5 \times 10^7$ CFU = Colony Forming Units, bei -70°C gelagert) werden aufgetaut und auf Eis gestellt
- 5 - Etwa 1ng des Ligationsansatzes aus b) wird dazu pipettiert und 20-30 Minuten auf Eis stehen gelassen, alle 5 Minuten vorsichtig mixen
- Hitzeschock: Ansatz 2 Minuten auf 42°C, zurück auf Eis
- 0,8ml LB-Medium (+ 0,02M MgSO_4 + 0,01M KCl) zugeben
- Ansatz 20-60 Minuten im Reagenzglas auf 37°C rollen
- 10 - 0,1 - 1 ml auf Agarplatte (mit Antibiotikum, z.B. Ampicillin) ausplattieren und Über Nacht auf 37°C

Die Transformation fungiert dabei als Selektionsverfahren auf erfolgreich klonierte Logomere und als Isolationsverfahren einzelner Logomere, weil jede Bakterienzelle genau ein Plasmid exprimiert.

15

Außer dem genannten Verfahren der Klonierung, das erfindungsgemäß bevorzugt ist, können die in Schritt V des erfindungsgemäßen Verfahrens erhaltenen informationstragenden Polymere (Logomere) mit folgenden Verfahren isoliert und vervielfältigt werden:

20 Isolation durch Hybridisierung und chromatographische Methoden

Hierbei wird das aus einer Symbolpolymerisation erhaltene Molekülgemisch. über ein Trägermaterial gegeben, an das Oligomere ("Ankersequenzen") gebunden sind, die ihrerseits an die Terminatoren der Logomere binden können. Die gebundenen Logomere werden dann einzeln aufgenommen und können nach Bedarf vervielfältigt werden.

25 Es sind hierzu verschiedene Verfahren einsetzbar:

- a) Affinitätschromatographie; dabei werden Ankersequenzen, die zu den überhängenden Enden der Terminatoren kompatible Enden besitzen, z.B. an eine Hydroxylapatit-Säule gebunden. Die aus Schritt V des erfindungsgemäßen Verfahrens erhaltenen Logomere werden über die Säule gegeben, wodurch die Logomere durch Hybridisierung an die Ankersequenzen binden können. Die solcherart gebundenen Logomere werden dann einzeln aufgenommen.
- 30 b) Ankersequenzen werden an eine Membran (z.B. Gene Screen Plus, DuPont, Biotechnology Systems; Hybond-N, Amersham Life Sciences) kovalent gebunden. Die Membran ist gerastert, d.h. in einzelne Felder unterteilt. Pro Feld wird genau eine Ankersequenz kovalent an die Membran gebunden. Danach werden die aus Schritt V des erfindungsgemäßen Verfahrens erhaltenen Logomere über die Membran gegeben und mit den Ankersequenzen hybridisiert. Die Membran wird daraufhin in die einzelnen Felder des Rasters zerschnitten. Die einzelnen Felder, die nun maximal ein an die jeweilige Ankersequenz gebundenes Logomer enthalten, können jetzt in separate Gefäße (z.B.
- 35 Eppendorf Tubes) überführt werden. Sodann können die Logomere durch Denaturieren von der Membran getrennt werden. Beim Denaturieren ist zu beachten, daß die
- 40

Denaturierungstemperatur so gewählt ist, daß sie hoch genug ist, um Logomer und Ankersequenz zu trennen, jedoch nicht so hoch, daß das Logomer selbst vollständig aufschmilzt.

5 Isolation durch Verdünnung

Bei der Isolation durch Verdünnung werden die Logomere, die als Gemisch aus einer Symbolpolymerisation erhalten werden, soweit verdünnt, daß in einem jeweiligen Zielvolumen statistisch genau ein Molekül enthalten ist. Dieses kann dann mit Hilfe der PCR aufgespürt und vervielfältigt werden. Die Verdünnung des Ausgangsvolumens in Zielvolumina kann dabei gleichzeitig geschehen, so daß ein Ausgangsvolumen auf n Zielvolumina verdünnt wird.

Dazu wird erfindungsgemäß wie folgt vorgegangen:

Zu einem, aus Schritt V des erfindungsgemäßen Verfahrens erhaltenen, Gemisch von Logomeren in einem Ausgangsvolumen V_a wird ein Verdünnungsfaktor ermittelt, der n in V_a enthaltene Logomere so auf n Zielvolumina V_z verdünnt, daß in jedem Zielvolumen V_z statistisch genau ein Logomer enthalten ist. Um den Verdünnungsfaktor zu ermitteln wird ein Aliquot des Ausgangsvolumens (z.B. $1\mu\text{l}$) in einer Verdünnungsreihe austitriert. Von jeder Verdünnung der Reihe wird dann ein Aliquot (z.B. $1\mu\text{l}$) als Template einer PCR-Reaktion verwendet, um zu ermitteln, in welchen Verdünnungen noch Logomere detektierbar sind. Erhält man solcherart die letzte Verdünnung, die noch Logomere enthält, und die Verdünnung, die schon keine mehr enthält, läßt sich ein ungefährer Verdünnungsfaktor angeben, der gerade noch ungefähr ein Logomer enthält. Gegebenenfalls kann dieser Verdünnungsfaktor durch Austitrieren der letzten Verdünnung, die noch Logomere enthält, genauer bestimmt werden. Dabei verwendet man entsprechend kleinere Verdünnungsschritte (verdünnt man z.B. in der ersten Verdünnungsreihe immer 1:10, so kann man in der 2. Verdünnungsreihe z.B. 1:2 verdünnen; das Verfahren der genaueren Bestimmung des Verdünnungsfaktors kann dabei prinzipiell mit beliebiger Genauigkeit angewendet werden).

Da die PCR der Verdünnungen dazu dient, festzustellen, ob sich überhaupt noch Logomere in der Verdünnung befinden, kann die PCR auf mindestens zweierlei Weise durchgeführt werden:

- 30 a) Als Primer dienen zwei gegenläufige, in den Terminatoren primende Primer, z.B. der 5'-Strang des Start-Terminators und der 3'-Strang des Ende-Terminators. Ansonsten sind die PCR-Bedingungen wie weiter unten unter *Vervielfältigung von Logomeren durch PCR* beschrieben.
- 35 b) Die PCR wird wie die zum Auslesen der Logomere durchgeführt, es reicht jedoch ein Ansatz mit einem Paar von Primern wie unter *Auslesen von Logomeren mittels PCR* im Folgenden beschrieben.

Zur Kontrolle werden die durch PCR erhaltenen DNA Fragmente durch Gelelektrophorese sichtbar gemacht. Dazu wird z.B. ein 2-4%iges Agarose-Gel (z.B. UltraPure™ von Gibco BRL, Life Technologies, Katalog Nr.: 15510-027) und als Molekulargewichtsstandard z.B. eine 50bp-Leiter (Gibco BRL, Life Technologies, Katalog Nr. 10416-014) verwendet. Das erhaltene Gel wird 5 Minuten in 0,001% Ethidiumbromid gefärbt und unter UV sichtbar gemacht.

Vervielfältigung von Logomeren

Die aus einer Symbolpolymerisation erhaltenen Logomere können – je nach ihrer beabsichtigten Verwendung – vervielfältigt werden. Dies ist beispielsweise für die
5 Visualisierung der gespeicherten Information oder die Weiterverwendung der Logomere als Marker zur Kennzeichnung von Stoffen und Gegenständen nötig.

Vervielfältigung von Logomeren durch PCR

10 Bei diesem Verfahren werden Logomere durch PCR vervielfältigt. Das zu vervielfältigende Logomer dient als Template, die benötigten Primer primen jeweils gegenläufig in den Terminatoren (entweder 5'-Start und 3'-Ende, oder 5'-Ende und 3'-Start), so daß das Logomer vollständig vervielfältigt wird. Alternativ können die Primer, sofern das zu vervielfältigende
15 Logomer von weiterer DNA umgeben ist, auch weiter außerhalb des Logomers primen.

Für die PCR-Ansätze, die in dem Fachmann bekannter Weise variiert werden können, werden z.B. folgende Reagentien und Bedingungen verwendet:
dNTPs (z.B. Pharmacia Biotech, Katalog Nr.: 27-2035-01/2/3), Taq-Polymerase (z.B. Gibco-BRL , Katalog Nr.n: 18038-034, 18038-042, 18038-067), PCR-Puffer, MgCl₂ (z.B. GIBCO-BRL, Katalog Nr.: 18067-017)
20

25

30

35

40

	Menge (μ l)
dNTP, 10mM	1
PCR-Puffer 10x	5
MgCl ₂ , 25mM	5
TAQ, 5u/ μ l	0,5
Primer 1, 10 μ M	1
Primer 2, 10 μ M	1
Template (Logomer)	1
H ₂ O	35,5
Gesamtvolumen	50

Als PCR-Programm kann beispielsweise folgendes Protokoll (Primer: 30-mer) verwendet werden:

Schritt	Aktion	Temperatur	Dauer	Gehe zu Schritt	Anzahl Wiederholungen
1	Denaturiere	95 °C	00:05:00		
2	Denaturiere	95 °C	00:00:30		
3	Annealing	68 °C	00:00:30		
4	Polymerisation	72 °C	00:00:30		
5	Gehe zu			2	29
6	Kühlen	4 °C	(beliebig)		
7	Ende				

- 5 Ein weiterer Gegenstand der Erfindung sind informationstragende Polymere, die gemäß den oben beschriebenen erfindungsgemäßen Verfahren erhältlich sind.

Auslesen der in Logomeren enthaltenen Information

10

Fertige Logomere können entweder mittels PCR (Polymerase Chain Reaction = Polymerasekettenreaktion), was erfindungsgemäß bevorzugt ist, oder mittels Restriktionsverdau gelesen werden. Dabei hat die PCR-Methode den Vorzug, bereits geringste Mengen DNA solcherart vervielfältigen zu können, daß die Logomere direkt nach

15

der PCR und einer Gel-Auftrennung gelesen werden können. Im Falle von binären Logomeren kann das durch die PCR-Methode auf einem Gel erhaltene Bandenmuster unmittelbar als Binärcode gelesen werden.

20

Ein weiterer Gegenstand der Erfindung ist daher ein Verfahren zum Auslesen von Information aus informationstragenden Polymeren, die wie oben beschrieben erhalten und/oder isoliert und vervielfältigt wurden, dadurch gekennzeichnet, daß man

- a) n das informationstragende Polymer enthaltende Lösungen mit jeweils einem Paar gegenläufiger Primer versetzt, wobei n für die Zahl der als Elongatoren in dem Polymer enthaltenen Oligomere steht;
- b) mindestens n-1 PCR-Ansätze durchführt, wobei n für die Zahl der als Elongatoren in dem Polymer enthaltenen Oligomere steht und jeweils ein Primer eines jeden Paares in dem dem Elongator gegenüberliegenden Terminator primt und der andere Primer in dem Elongator selbst primt;
- c) die aus der PCR erhaltenen Polymer-Fragmente nach ihrer Länge durch Elektrophorese auftrennt und
- d) das aus der Elektrophorese erhaltene Muster optisch ausliest.
- Das Verfahren kann auch unter Verwendung verschiedenfarbiger fluoreszenzmarkierter Primer oder Nukleotide durchgeführt werden. Dadurch wird es möglich, mehrere Ansätze verschiedenfarbig zu markieren und mehrere Ansätze in einer Spur gelelektrophoretisch zu lesen. Überdies kann der Ausleseprozeß dadurch mit modernen Sequenziermaschinen (z.B. von ABI erhältlich) automatisiert werden.

Auslesen von Logomeren mittels PCR

- Um die in Logomeren enthaltenen Informationen sichtbar zu machen, können die Logomere mithilfe der PCR ausgelesen werden.
- Die Methode des Auslesens binärer Muster mit PCR wurde als *DNA Typing* bereits von [Jeffreys, Minisatellite repeat coding as a digital approach to DNA typing, *Nature*, **354**, 204-209, (1991)] beschrieben. Die dort beschriebene Methode wird hier in einer abgewandelten Form als *Auslese-PCR* zum Auslesen der Logomere benutzt. Unterschiede zu der in [Jeffreys, Minisatellite repeat coding as a digital approach to DNA typing, *Nature*, **354**, 204-209, (1991)] beschriebenen Methode bestehen darin, daß hier synthetisch erzeugte Templates verwendet werden, die nicht notwendig binäre Muster enthalten. Außerdem werden PCR- und Gel-Elektrophorese-Bedingungen so gewählt, daß das Ergebnis direkt vom Gel gelesen werden kann, ohne daß die als Bandenmuster erhaltenen Signale zusätzlich durch Blotting verstärkt werden müssen.

- a) PCR
- Für das Auslesen eines Logomers werden für n Elongatoren der zugrundeliegenden Grammatik mindestens n - 1, meistens aber n PCR-Ansätze gebraucht. Jeder PCR-Ansatz enthält das zu lesende Logomer als Template und außerdem ein Paar gegenläufiger Primer, von denen einer im Elongator, der andere im, dem Elongator gegenüberliegenden, Terminator primt (z.B. 5'-"Start" und 3'-O, siehe Abbildung 5). Für die PCR kann jeder handelsübliche Thermozykler (z.B. PTC-100, MJ Research) verwendet werden. Als zweckmäßig für die Länge der Primer haben sich 20-30bp erwiesen, jedoch sind auch andere Längen verwendbar. Je nach Primerlänge muß die Annealing-Temperatur entsprechend gewählt

werden (kann z.B. mithilfe des Programmes Oligo 5.0 bestimmt werden). Zweckmäßige Annealingtemperaturen sind für 20-mere ca. 55°C, für 30-mere etwa 65°C bis 74 °C.

Für die PCR-Ansätze werden z.B. folgende Reagentien und Protokolle verwendet, die in dem Fachmann bekannter Weise variiert werden können:

- 5 dNTPs (Pharmacia Biotech, Katalog Nr.: 27-2035-01/2/3), Taq-Polymerase (Gibco-BRL, Katalog Nr.: 18038-034, 18038-042, 18038-067), PCR-Puffer, MgCl₂ (GIBCO-BRL, Katalog Nr.: 18067-017)

	Menge (µl)
dNTP, 10mM	1
PCR-Puffer 10x	5
MgCl ₂ , 25mM	5
TAQ, 5u/µl	0,5
Primer 1, 10µM	1
Primer 2, 10µM	1
Template (Logomer)	1
H ₂ O	35,5
Gesamtvolumen	50

- 10 Um genügend DNA zu erhalten, so daß das erhaltene Bandenmuster nach der Gelelektrophorese direkt vom Gel gelesen werden kann, kann jeder PCR-Ansatz m mal angesetzt werden. Z.B. kann m = 4 sein.

Als PCR-Programm kann folgendes Protokoll (Primer: 30-mere) verwendet werden:

Schritt	Aktion	Temperatur	Dauer	Gehe zu	Anzahl
				Schritt	Wiederholungen
1	Denaturiere	95 °C	00:05:00		
2	Denaturiere	95 °C	00:00:30		
3	Annealing	69,5 °C	00:00:30		
4	Polymerisation	72 °C	00:00:30		
5	Gehe zu			2	29
6	Kühlen	4 °C	(beliebig)		
7	Ende				

- 15 Nach durchgeführter PCR ist es für die Lesbarkeit der Bandenmuster der nachfolgenden Gelelektrophorese evtl. zweckmäßig, möglichst viel amplifizierte DNA in möglichst geringem Volumen zu halten, das dann auf ein Gel geladen werden kann. Dazu können die Volumina auf geeignete Weise (z.B. mit einer SpeedVac, z.B. SpeedVac Concentrator, Savant) eingeengt werden.
- 20 b) Gelelektrophorese

Die für jeden Elongator aus der PCR erhaltenen DNA Fragmente haben verschiedene Länge. Werden sie durch Elektrophorese aufgetrennt, so ergeben die unterschiedlichen Längenfragmente ein spezifisches Muster, anhand dessen sich das zu lesende Logomer identifizieren läßt (siehe Abbildung 5 und Abbildung 6). Für die Gelelektrophorese kann ein 4% Agarose-Gel (z.B. UltraPure™ von Gibco BRL, Life Technologies, Katalog Nr.: 15510-027), als Molekulargewichtsstandard z.B. eine 50bp-Leiter (Gibco BRL, Life Technologies, Katalog Nr.: 10416-014) verwendet werden. Die Auftrennung der DNA erfolgt in einer Gelkammer in geeigneter Weise, z.B. 1:45 Stunden bei 60V. Hier sind die Parameter jedoch relativ frei wählbar, insbesondere kann die Gel-Laufzeit weiter verkürzt werden.

Das erhaltene Gel wird 5 Minuten in 0,001% Ethidiumbromid gefärbt und unter UV sichtbar gemacht.

Ein Beispiel für ein solcherart erhaltenes Gel findet sich exemplarisch unter Abbildung 6.

Auslesen von Logomeren durch Restriktionsverdau

Logomere können durch Restriktionsverdau ausgelesen werden. Dazu müssen die Elongatoren so konstruiert sein, daß sie asymmetrisch versetzte Restriktionsschnittstellen tragen. Jeder Elongator trägt dabei eine spezifische Restriktionsschnittstelle. Beispielsweise kann das Restriktionsenzym EcoRV (z.B. NEB, Katalog Nr. #195S) für 0-Elongatoren und SmaI (z.B. NEB, Katalog Nr. #141S) für 1-Elongatoren verwendet werden.

Zum Lesen wird das zu lesende Logomer in verschiedenen Restriktionsansätzen geschnitten. Dazu wird pro Elongator ein Restriktionsansatz mit dem jeweiligen Restriktionsenzym gebildet. Die aus der Restriktion erhaltenen DNA Fragmente sind von verschiedener Länge. Werden sie durch Elektrophorese aufgetrennt, so ergeben die unterschiedlichen Längenfragmente ein spezifisches Muster, anhand dessen sich das zu lesende Logomer identifizieren läßt.

Das folgende Beispiel zeigt die DNA-Fragmentlängen aller binären Logomere mit 4 Bit ohne Terminatoren mit Elongatorlänge = 30bp; Restriktionsschnittstelle pro Elongator nach 10bp.

Binärzahl		0	1
5	0000	10, 40, 40, 40, 30	160
	0001	10, 40, 40, 70	130, 30
	0010	10, 40, 80, 30	90, 70
	0011	10, 40, 110	90, 40, 30
	0100	10, 80, 40, 30	50, 110
10	0101	10, 80, 70	50, 80, 30
	0110	10, 120, 30	50, 40, 70
	0111	10, 150	50, 40, 40, 30
	1000	50, 40, 40, 30	10, 150
	1001	50, 40, 70	10, 120, 30
15	1010	50, 80, 30	10, 80, 70
	1011	50, 110	10, 80, 40, 30
	1100	90, 40, 30	10, 40, 110
	1101	90, 70	10, 40, 80, 30
	1110	130, 30	10, 40, 40, 70
	1111	160	10, 40, 40, 40, 30

Im obigen Beispiel (4 Bits) erkennt man, daß die Längenmuster der Zahlen 0010 und 0100 bei Restriktion des 0-Bits nicht eindeutig ist. 0010 und 0100 lassen sich jedoch anhand der Längenmuster bei Restriktion des 1-Bits unterscheiden.

Im Einzelnen wurde zur Durchführung des obigen Beispiels folgendermaßen vorgegangen, wobei sich die Protokolle gemäß den Kenntnissen des Fachmanns variieren lassen:

a) DNA Extraktion

Eine z.B. durch *Vereinzelung und Vervielfältigung von Logomeren durch Klonierung* erhaltene Bakterienkolonie wird zum Animpfen von 2-5ml einer 37°C Übernachtskultur verwendet (z.B. LB-Medium mit 10µg/l Ampicillin wie in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)] beschrieben). Als Klonierungsvektor können hier z.B. pGEM-T Easy (Promega, Katalog Nr. A1360) und binäre Logomere auch ohne Terminatoren verwendet werden. Aus der Übernachtskultur wird die das Logomer enthaltende Plasmid-DNA isoliert (z.B. mithilfe des Qiagen Plasmid Miniprep Kit, Qiagen, Katalog Nr. 12123, oder durch alkalische Lyse wie in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)] beschrieben).

b) Restriktion

Die aus a) erhaltene Plasmid DNA wird für n Elongatoren in n Restriktionsansätzen geschnitten. Jeder Restriktionsansatz enthält ein Restriktionsenzym, das in einem spezifischen Elongator schneidet und ein Restriktionsenzym, das das Logomer aus der Klonierungssite ausschneidet. Im verwendeten Beispiel wurden folgende Ansätze verwendet:

	Name	Menge	Volumen (µl)
Plasmid	Logomer in pGEM-T	500ng/µl	6
DNA	Easy (s.o.)		
Puffer	NEB2	10x	1

36

BSA		10x	1
Enzym 1	Eco RV	10u/ μ l	1
Enzym 2	Eco RI	10u/ μ l	1
Total			10

	Name	Menge	Volumen (μ l)
Plasmid	Logomer in pGEM-T	500ng/ μ l	6
DNA	Easy (s.o.)		
Puffer	NEB4	10x	1
BSA		10x	0
Enzym 1	Sma I	10u/ μ l	1
Enzym 2	Eco RI	10u/ μ l	1
H ₂ O			1
Total			10

Die angegebenen Ansätze werden 1 Stunde auf 37°C inkubiert.

c) Gelelektrophorese

- 5 Die aus b) erhaltenen DNA Fragmente werden elektrophoretisch aufgetrennt (z.B. 4% Agarose UltraPure™ von Gibco BRL, Life Technologies, Katalog Nr.: 15510-027), mit Ethidiumbromid (0,001%) gefärbt und unter UV sichtbar gemacht (siehe Abbildung 8).
- 10 Die oben beschriebenen Verfahren ermöglichen beispielsweise die Implementierung binärer Logomere:

Binäre Logomere werden mit Grammatiken erzeugt, in denen gilt:

$\Sigma := \{0, 1, s_0, s_1, s_2, \dots, s_{n-1}, s_n, e_0, e_1, e_2, \dots, e_{m-1}, e_m\}$ mit $n, m \in \mathbb{N}$ und $n, m \geq 0$.

- 15 Sie ermöglichen die einfachste und universellste Darstellung von Daten, wie sie auch von herkömmlichen Datenverarbeitungsmaschinen eingesetzt wird. Binäre Logomere sind Resultat einer Bitpolymerisation. Da abhängig von der jeweils gewählten Grammatik nahezu beliebige Symbole und Regeln kodiert werden können, können damit ebenso beliebige Daten und Datentypen erzeugt werden.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Implementierung von Zeichenalphabeten:

Logomere können eingesetzt werden, um die Zeichen eines gewählten Alphabetes zu kodieren. Die Zeichenlänge kann dabei unterschiedlich gewählt werden, sinnvoll ist es jedoch, die auch für herkömmliche Datenverarbeitungsanlagen benutzten Zeichenlängen (Halbbyte, 1-Byte, 2-Byte, 4-Byte, 8-Byte usw.) zu verwenden. Ebenso ist die Konvention für die Interpretation der dargestellten Zeichen wahlfrei. Die Zeichen können Zahlen, Buchstaben, alphanumerische Zeichen oder beliebige andere Datenstrukturen sein.

10

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Implementierung eines 1-Byte Alphabetes:

Es sei eine Grammatik $G = (\Sigma, V, R, S)$ mit Terminalalphabet $\Sigma := \{0, 1, s, e\}$, Variablenmenge $V := \{S_0, S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8\}$, Startsymbol S und Regelmenge

15

$R :=$

{

20

$S := sS_0$

$S_0 \rightarrow 0S_1$

$S_0 \rightarrow 1S_1$

$S_1 \rightarrow 0S_2$

$S_1 \rightarrow 1S_2$

$S_2 \rightarrow 0S_3$

$S_2 \rightarrow 1S_3$

25

$S_3 \rightarrow 0S_4$

$S_3 \rightarrow 1S_4$

$S_4 \rightarrow 0S_5$

$S_4 \rightarrow 1S_5$

$S_5 \rightarrow 0S_6$

30

$S_5 \rightarrow 1S_6$

$S_6 \rightarrow 0S_7$

$S_6 \rightarrow 1S_7$

$S_7 \rightarrow 0S_8$

$S_7 \rightarrow 1S_8$

35

$S_8 \rightarrow e$

}

wobei:

$s := \text{Start}$

$e := \text{Ende.}$

40

Beispiel:

nach den Regeln der angegebenen Grammatik können alle 8-Bit Zeichen erzeugt werden:

z.B.: Erzeugung der 8-Bit 0, 00000000:

5 S-> sS₀ -> s0S₁ -> s00S₂ -> s000S₃ -> s0000S₄ -> s00000S₅ -> s000000S₆ -> s0000000S₇ -> s00000000S₈ -> s00000000e

z.B.: Erzeugung der 8-Bit 255, 11111111:

S-> sS₁ -> s1S₁ -> s11S₂ -> s111S₃ -> s1111S₄ -> s11111S₅ -> s111111S₆ -> s1111111S₇ -> s11111111S₈ -> s11111111e

z.B.: Erzeugung der 8-Bit 85, 01010101:

10 S-> sS₀ -> s0S₁ -> s01S₂ -> s010S₃ -> s0101S₄ -> s01010S₅ -> s010101S₆ -> s0101010S₇ -> s01010101S₈ -> s01010101e

Die für die Implementierung in vitro benötigten Sequenzen werden nach dem NFR-Verfahren, wie in den erfindungsgemäßen Schritten II und III beschrieben, hergestellt. Algomere und Logomere werden, wie in den erfindungsgemäßen Schritten IV - V beschrieben, hergestellt und die erhaltenen Logomere vorzugsweise mittels des unter *Vereinzelung und Vervielfältigung von Logomeren durch Klonierung* beschriebenen Verfahrens isoliert und vervielfältigt. Die solcherart erzeugten alphanumerischen Zeichen können jetzt für verschiedene technische Zwecke (z.B. die Markierung und Identifizierung von Stoffen) eingesetzt werden und, evtl. nach zusätzlichen technischen Schritten, gemäß dem erfindungsgemäßen Verfahren zum Auslesen von Information aus informationstragenden Polymeren (wie oben beschrieben) ausgelesen werden.

Da die hier erzeugten Binärmuster wahlfrei als Daten und Symbole gelesen werden können, können sie z.B. als alphanumerische Zeichen interpretiert werden. Werden sie als Zahlen interpretiert, so erzeugt die angegebene Grammatik 8-Bit Zufallszahlen.

Aus einer genügend großen Menge zufälliger 8-Bit Binärmuster können alle 8-Bit Muster isoliert und als Bibliothek (z.B. zur Darstellung aller alphanumerischen Zeichen) angelegt werden.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Darstellung von Zeichenketten:

Da Algomere so hergestellt werden können, daß sich unterschiedlich lange Logomere erzeugen lassen, ist es möglich, auch Zeichenketten darzustellen. Aus praktischen Gründen ist es jedoch zweckmäßig, nur wenige Zeichen pro Logomer vorzusehen. Z.B. kann ein einzelnes Logomer 4 Bit (Halbbyte) oder 8 Bit (1 Byte) enthalten. (Falls mehr Bit erforderlich sind, sollten zweckmäßigerweise immer Vielfache von 8 Bit = 1 Byte verwendet werden.) In Binärdarstellung können dann beliebige alphanumerische Zeichen in beliebiger Kodierung (z.B. ASCII, ANSI, ISO 8859-x, Unicode) dargestellt werden. Dabei kann sich ein Zeichen auch über mehrere Logomere erstrecken (Bsp.: Halbbyte-Darstellung, bei der sich jeweils ein

1-Byte Zeichen über zwei Logomere erstreckt, oder Unicode, bei der sich ein 16-Bit Zeichen über zwei 8-Bit oder vier 4-Bit Logomere erstreckt). Für die Darstellung von Zeichenketten ist es dann nötig, die einzelnen Zeichen mit Positionsinformation zu versehen. Dazu reicht es, in den jeweiligen Grammatiken verschiedene Terminatoren (z.B. verschiedene "Start"-

- 5 Terminatoren) zu verwenden, deren Sequenzen jeweils die Positionsinformation repräsentieren.

Beispiel: Es sei eine Grammatik $G = (\Sigma, V, R, S)$ mit Terminalalphabet $\Sigma := \{0, 1, e, s_0, s_1, s_2, \dots, s_{n-1}, s_n\}$, Variablenmenge $V := \{S_0, S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8\}$, Startsymbol S und Regelmenge

10 $R :=$

{

$S := s_i S_0$

$S_0 \rightarrow 0 S_1$

$S_0 \rightarrow 1 S_1$

15 $S_1 \rightarrow 0 S_2$

$S_1 \rightarrow 1 S_2$

$S_2 \rightarrow 0 S_3$

$S_2 \rightarrow 1 S_3$

$S_3 \rightarrow 0 S_4$

20 $S_3 \rightarrow 1 S_4$

$S_4 \rightarrow 0 S_5$

$S_4 \rightarrow 1 S_5$

$S_5 \rightarrow 0 S_6$

$S_5 \rightarrow 1 S_6$

25 $S_6 \rightarrow 0 S_7$

$S_6 \rightarrow 1 S_7$

$S_7 \rightarrow 0 S_8$

$S_7 \rightarrow 1 S_8$

$S_8 \rightarrow e$

30 }

wobei:

$s_i := \text{Start}$, mit $i = \{0, \dots, n\}$

$e := \text{Ende}$.

- 35 Das mit dieser Grammatik erzeugte 1-Byte Alphabet reicht z.B. für alle alphanumerischen Zeichen des ASCII, ANSI oder ISO 8859-x Standards. Für eine Zeichenkette der Länge n werden n verschiedene Terminatoren benötigt, so daß die mit dieser Grammatik erzeugten Zeichenketten wie folgt aufgebaut sind:

$s_0 x e, s_1 x e, \dots, s_{n-1} x e, s_n x e$

- 40 wobei x eine beliebige Binärdarstellung mit, in diesem Fall, 8 Bit ist.

Es bedeutet dann:

s_0x_e := Zeichen x an Position 0 (0.tes Zeichen)

s_1x_e := Zeichen x an Position 1 (1.tes Zeichen)

s_2x_e := Zeichen x an Position 2 (2.tes Zeichen)

5 usw.

Mit der Grammatik läßt sich beispielsweise die Zeichenkette "Elisabeth" im ASCII-Code darstellen:

$s_001000101e$, $s_101101100e$, $s_201101001e$, $s_301110011e$, $s_401100001e$, $s_501100010e$,
 $s_601100101e$, $s_701110100e$, $s_801101000e$

10 Alternativ kann auch $\Sigma := \{0, 1, s, e_0, e_1, e_2, \dots, e_{n-1}, e_n\}$ gewählt werden, wodurch die Zeichenkette "Elisabeth" im ASCII-Code als:

$s01000101e_0$, $s01101100e_1$, $s01101001e_2$, $s01110011e_3$, $s01100001e_4$, $s01100010e_5$,
 $s01100101e_6$, $s01110100e_7$, $s01101000e_8$

dargestellt wird. Eine weitere Möglichkeit ist $\Sigma := \{0, 1, s_0, s_1, s_2, \dots, s_{n-1}, s_n, e_0, e_1, e_2, \dots, e_{n-1},$

15 $e_n\}$, womit die Zeichenkette "Elisabeth" im ASCII-Code als:

$s_001000101e_0$, $s_101101100e_1$, $s_201101001e_2$, $s_301110011e_3$, $s_401100001e_4$, $s_501100010e_5$,
 $s_601100101e_6$, $s_701110100e_7$, $s_801101000e_8$

dargestellt wird.

In Halbbyte-Darstellung mit Positionsinformation würde die Zeichenkette "Elisabeth" im ASCII-

20 Code dargestellt als:

s_00100e , s_10101e , s_20110e , s_31100e , s_40110e , s_51001e , s_60111e , s_70011e , s_80110e , s_90001e ,
 $s_{10}0110e$, $s_{11}0010e$, $s_{12}0110e$, $s_{13}0101e$, $s_{14}0111e$, $s_{15}0100e$, $s_{16}0110e$, $s_{17}1000e$

Aufgrund der Positionsinformation, die durch die Sequenz der Terminatoren repräsentiert wird, können die einzelnen Zeichen völlig unabhängig voneinander verarbeitet werden und z.B. getrennt voneinander gelesen werden.

Mit einem solchen Alphabet ist die Darstellung beliebiger Datentypen möglich. Z.B. können z.B. jeweils zwei Byte (0. + 1., 2. + 3., ..., n. + n+1.) zu einem 2-Bytecode (z.B. Unicode) zusammengefaßt werden. Alphanumerische Zeichenketten können verwendet werden, um

30 beliebige Bezeichner (Namen, Zahlen, Datum usw.) darzustellen.

Die für die Implementierung in vitro benötigten Sequenzen werden nach dem NFR-Verfahren, wie in den erfindungsgemäßen Schritten II und III beschrieben, hergestellt. Algomere und Logomere werden, wie in den erfindungsgemäßen Schritten IV - V beschrieben, hergestellt, die erhaltenen Logomere vorzugsweise mittels des unter *Vereinzelung und Vervielfältigung von Logomeren durch Klonierung* beschriebenen Verfahrens isoliert und vervielfältigt. Die solcherart erzeugten alphanumerischen Zeichen und Zeichenketten können jetzt für verschiedene technische Zwecke (z.B. die Markierung und Identifizierung von Stoffen) eingesetzt werden und, evtl. nach zusätzlichen technischen Schritten, gemäß dem

40 erfindungsgemäßen Verfahren zum Auslesen von Information aus informationstragenden Polymeren (wie oben beschrieben) ausgelesen werden.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Implementierung eines Zufallszahlengenerators:

Für bestimmte Probleme der Informatik (z.B. Simulationen) und der Mathematik werden
 5 Zufallszahlen benötigt. Diese werden im allgemeinen rechnergestützt durch Algorithmen erzeugt. Da diese Algorithmen jedoch deterministisch sind, handelt es sich bei den erzeugten Zufallszahlen nur um Pseudo-Zufallszahlen. Überdies sind die erzeugten Zahlenreihen aufgrund der unterschiedlichen Güte der Algorithmen unterschiedlich gut randomisiert.

Für manche Anwendungen werden daher echte Zufallszahlen benötigt. Ist dies der Fall, muß
 10 ein physikalischer Prozeß in die Erzeugung von Zufallszahlen einbezogen werden. Ein solcher Prozeß ist etwa das Rauschen der Soundkarte in einem PC, das von entsprechender Software verarbeitet wird.

Mithilfe der hier beschriebenen Verfahren kann ein echter Zufallszahlengenerator implementiert werden, der Zufallszahlen sehr viel schneller erzeugt, als es mit herkömmlichen
 15 Verfahren möglich ist.

Der Zufallszahlengenerator wird mit folgender Grammatik implementiert (Grammatik für binäre Zufallszahlen beliebiger Länge):

$G = (\Sigma, V, R, S)$ mit $\Sigma := \{0, 1, s, e\}$, $V := \{S\}$,

$R :=$

20 {
 $S := sA$
 $A \rightarrow 0A$
 $A \rightarrow 1A$
 $A \rightarrow e$

25 }
 wobei

$s := \text{Start}$

$e := \text{Ende}$

Mit dieser Grammatik können Wörter über dem oben angegebenen Terminalalphabet

30 $\Sigma := \{0, 1, s, e\}$

gebildet werden. Alle Wörter der Sprache L beginnen mit „s“ und hören mit „e“ auf, dazwischen befindet sich eine zufällige Anzahl (auch 0) zufälliger Bits („0“ und „1“).

Beispiel: Wörter aus der Sprache L , die nach R gebildet werden:

35 $S \rightarrow sA \rightarrow se$

$S \rightarrow sA \rightarrow s1A \rightarrow s1e$

$S \rightarrow sA \rightarrow s0A \rightarrow s00A \rightarrow s001A \rightarrow s0010A \rightarrow s0010e$

$S \rightarrow sA \rightarrow s1A \rightarrow s10A \rightarrow s100A \rightarrow s1000A \rightarrow s10000A \rightarrow s100000A \rightarrow s100000e$

Die als Wörter der Sprache L erzeugten Binärmuster können als Buchstaben, Zahlen, alphanumerische Zeichen oder Zeichenketten gelesen werden. Liest man sie als die Binärdarstellung von Zahlen, so sind diese Binärmuster in Dezimaldarstellung:

Ø (leer)

5 1
2
32

und können als Zufallszahlen verwendet werden. Zur in vitro Implementierung eignen sich prinzipiell beliebige Sequenzen, solange sie eindeutig und zueinander maximal unähnlich sind.

10 Z.B. können folgende Algomere verwendet werden:

sA:

```
agctttatatctccatttgccttagtgaag
aatatagaggtaaacgggatcacttcaacc
```

A->0A:

```
ttggcgagatatcaacgccaccccttgctt
gctctatagttgcggtggggaacgaaaacc
```

15 A->1A:

```
ttggcgacccgggaaacaactattgctagt
gctgggccctttgttgataacgatcaaacc
```

A->e:

```
ttggtgcgggagttggaagcaactacgatg
acgccctcaaccttcggttgatgctacctag
```

Die dazu benötigten Sequenzen sind handelsüblich. Sie können z.B. bestellt werden als 20 40nmol, PAGE gereinigt (ARK-Scientific, Darmstadt) und z.B. 100µM in Aqua dest. aufgenommen werden (empfohlene Lagerung bei -20°C). Aus den Sequenzen werden, wie im erfindungsgemäßen Schritt IV beschrieben, Algomere hergestellt. Diese Algomere werden, wie im erfindungsgemäßen Schritt V beschrieben, zu Logomeren polymerisiert und können gemäß dem erfindungsgemäßen Verfahren zum Auslesen von Information aus 25 informationstragenden Polymeren (wie oben beschrieben) ausgelesen werden. Derart erzeugte Zufallszahlen sind in Abbildung 6 zu sehen.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Verschlüsselung 30 von Logomeren:

Die in Logomeren enthaltene Information kann verschlüsselt werden. Dazu fungiert mindestens einer der für die Auslese-PCR benötigten Primer als geheimer Schlüssel (siehe Abbildung 11). Bevorzugt ist dies einer der in den Terminatoren primenden Primer. Ist die

Sequenz dieses Primers (Schlüsselsequenz) und der Primer selbst nur autorisierten Zugreifern bekannt, so ist die in den derart verschlüsselten Logomeren enthaltene Information auch nur autorisierten Zugriffen zugänglich.

Um ein Auslesen der in den Logomeren enthaltenen Information ohne diesen Schlüssel zu verhindern, müssen weitere Vorkehrungen getroffen werden. Versuche, die Verschlüsselung zu brechen, können darauf beruhen, über die Vervielfältigung nicht-geheimer Sequenzen auch die Logomere zu lesen. Sofern bakterielle Vektoren zur Klonierung von Logomeren verwendet werden, könnte ein potentieller Angreifer z.B. versuchen, die gesamte Klonierungssite mit PCR zu vervielfältigen und dann mit Sequenzierung zu lesen oder über die zur Selektion benötigten Resistenzgene in die Klonierungssite "hineinzulesen". Auch könnten die Elongatoren selbst als Ansatzpunkte für ein PCR-basiertes Lesen der geheimen Sequenz(en) dienen.

Solchen Angriffsversuchen ist gemein, daß sie über nicht-geheime Sequenzen versuchen könnten, die Schlüsselsequenz zu entschlüsseln und darüber die verschlüsselte Sequenz zu lesen. Eine Abwehr derartiger Angriffsversuche kann dadurch erfolgen, daß zum Informationstransport über Logomere immer nur vollständig geheime Sequenzen verwendet werden. Jedoch ist dies evtl. zu aufwendig und kostenträchtig. Stattdessen oder ergänzend dazu können die informationstragenden Logomere mit Sequenzen (Scheinsequenzen) versetzt werden, die dieselben potentiellen Angriffspunkte (z.B. Bits, Resistenzgene) enthalten, jedoch jeweils andere Schlüsselsequenzen. Durch diese Scheinsequenzen laufen potentielle Angriffsversuche ins Leere, weil für den nicht-autorisierten Zugreifer alle Einzelsequenzen ununterscheidbar und dadurch die verschlüsselten Informationen verborgen sind. Mit je mehr Scheinsequenzen ein Logomer versetzt ist, um so schwieriger wird es, die Schlüsselsequenz zu brechen.

Das Verfahren entspricht einer molekularen Steganografie und wird exemplarisch anhand der Verschlüsselung von Zeichen eines 1-Byte Alphabetes gezeigt:

Es sei die verwendete Grammatik $G = (\Sigma, V, R, S)$ mit Terminalalphabet $\Sigma := \{0, 1, s_{key}, s_0, s_1, s_2, \dots, s_{f-1}, s_f, e\}$, wobei $f \in \mathbb{N}$, $f \geq 1$. s_{key} ist der geheime Schlüssel. Für eine Schlüssellänge von l ist dann die maximale theoretische Verschlüsselung $Key_{max} = a^l$, die maximal verwendbare Verschlüsselung $Key_{eff} = a^{l-d}$, wobei d die Anzahl der Basen angibt, in der sich ein Scheinschlüssel von dem richtigen Schlüssel unterscheiden muß, damit in der Auslese-PCR kein Scheinschlüssel die richtige Sequenz lesen und umgekehrt der richtige Schlüssel kein Scheinlogomer (sondern nur das Ziellogomer) auslesen kann. Für hochspezifische PCR Konditionen kann $d \geq 1$ werden. Weiterhin gibt $Key_{imp} = \text{Anzahl Scheinschlüssel} + 1$ die tatsächlich verwendete Verschlüsselung und $Key_{min} = 1 + x$ die minimale Verschlüsselung an. Dabei ist x die zu einer minimalen Verschlüsselung nötige Anzahl von Scheinlogomeren. Diese kann im Falle der Verwendung von n Logomeren als Informationsträger größer als n gewählt werden.

Für Zeichenketten ergibt sich automatisch eine zusätzliche Verschlüsselung, weil hier die für einen potentiellen Angreifer unbekannte Reihenfolge der Zeichen zusätzlich verschlüsselnd wirkt.

Beispiel: Für eine Grammatik mit Zeichenketten von n 1-Byte Zeichen wird folgende Grammatik G verwendet:

Es sei $G = (\Sigma, V, R, S)$ mit Terminalalphabet $\Sigma := \{0, 1, S_{key0}, S_{key1}, S_{key2}, \dots, S_{keyn-1}, S_{keyn}, S_0, S_1, S_2, \dots, S_{i-1}, S_i, e\}$, wobei $n, f \in \mathbb{N}$, $n \geq 1$. n gibt dabei die Anzahl der verwendeten echten Schlüssel, f die Anzahl der verwendeten Scheinschlüssel an; die verwendete Variablenmenge V und die verwendete Regelmenge R sind dabei wahlfrei und abhängig von der zu kodierenden Information.

Grammatiken mit anderer Zeichenkodierung funktionieren analog.

Die für die Implementierung in vitro benötigten Sequenzen werden nach dem NFR-Verfahren, wie in den erfindungsgemäßen Schritten II und III beschrieben, hergestellt. Algomere und Logomere werden, wie in den erfindungsgemäßen Schritten IV - V beschrieben, hergestellt, die erhaltenen Logomere vorzugsweise mittels des unter *Vereinzelung und Vervielfältigung von Logomeren durch Klonierung* beschriebenen Verfahrens isoliert und vervielfältigt. Die solcherart erzeugten Logomere können, wie bereits als Auslese-PCR beschrieben, ausgelesen werden, wobei - wie beschrieben - einer der zum Auslesen benötigten Primer, bevorzugt einer der in den Terminatoren primenden Primer, als geheimer Schlüssel fungiert. Mit der oben beschriebenen Grammatik werden zusätzlich zu den informationstragenden Logomeren „Scheinsequenzen“ („Scheinlogomere“) erzeugt, die einen nicht-autorisierten Auslese-Zugriff auf die in den Logomeren enthaltenen Informationen verhindern sollen. Das Verfahren kann für alle Anwendungen von Logomeren benutzt werden und hat u.a. den Vorzug, durch die Spezifität des als geheimer Schlüssel benutzten Primers sehr effektiv zu sein.

Basierend auf der beschriebenen Methode können symmetrische und asymmetrische Verschlüsselungsverfahren implementiert werden. Für ein symmetrisches Verschlüsselungsverfahren reicht es, daß sich die Teilnehmer A und B einer Kommunikation auf eine geheimgehaltene Schlüsselsequenz einigen. Teilnehmer B verschlüsselt eine als Logomer gespeicherte Nachricht an A, indem B Logomere nur mit den geheimen Terminatoren herstellt und seine Nachricht mit zusätzlichen Logomeren versetzt, die andere Terminatoren tragen. Nur A ist es dann möglich, das zum Auslesen benötigte Primerpaar bereitzustellen, da nur A die benötigte Terminatorsequenz kennt.

Ein asymmetrisches Verschlüsselungsverfahren erfordert dagegen zusätzlichen Aufwand, z.B. den Einsatz von Molekülen mit irregulären Formen (siehe z.B. Abbildung 12 und Abbildung 13): Will ein Kommunikationsteilnehmer B eine verschlüsselte Nachricht an A schicken, so bekommt er von A einen öffentlichen Schlüssel, den er zur Verschlüsselung der als Logomer repräsentierten Nachricht verwenden kann. Der öffentliche Schlüssel von A besteht aus einem Paar von Terminatoren und einem Pool zusätzlicher Scheinterminatoren mit ähnlichen Eigenschaften, jedoch abweichenden Sequenzen und anderen 3'-Überhangsequenzen. Von den echten Terminatoren ist lediglich die 3' Überhangsequenz bekannt, mit der sie an die öffentlich bekannten - Elongatoren verknüpft werden kann. Zur Verschlüsselung einer

45

Nachricht erzeugt B Logomere, wobei er in der Symbolpolymerisationsreaktion den von A öffentlich bereitgestellten Schlüssel (enthaltend Terminatoren und Scheinterminatoren) verwendet. Die Art und Weise der Terminatoren gewährleistet nun, daß lediglich A in der Lage ist, die solcherart verschlüsselte Nachricht wieder zu lesen: Da nur A die tatsächliche Sequenz der echten Terminatoren kennt, ist nur A in der Lage, die als Logomere verschlüsselte Nachricht über Auslese-PCR wieder zu lesen.

Weil der öffentliche Schlüssel potentiell über die bekannte Überhangsequenz zu den Elongatoren angreifbar ist (weil die Scheinterminatoren diese Überhangsequenz nicht besitzen dürfen, kann ein potentieller Angreifer eine beliebige Sequenz an die echten Terminatoren knüpfen, wodurch die Sequenz der Terminatoren identifiziert werden kann), werden als Terminatoren Moleküle auf der Basis irregulärer Formen, z. B. auf der Basis der in Abbildung 12 gezeigten Y-förmigen Moleküle verwendet. Die gezeigten Y-förmigen Moleküle lassen sich zu weiter verzweigten baumartigen Molekülen zusammensetzen. Die Wurzel eines solchen als Baum realisierten Terminatormoleküls enthält dann die zu den Elongatoren kompatible Überhangsequenz, während nur eine der Äste des Baumes die echte Terminatorsequenz enthält. Da die verschlüsselte Nachricht wie oben beschrieben zusätzlich mit Scheinlogomeren versetzt ist, kann dann nur A, der die echte Terminatorsequenz kennt, aus der Menge der Moleküle die richtige Nachricht durch Verknüpfen mit dem kompatiblen Vektor herausfiltern, wohingegen ein potentieller Angreifer ohne Kenntnis der echten Terminatorsequenz dies nicht kann.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Verschlüsselung mit Echtzeit-Identifizierung:

25

Ohne das Auslesen der in verschlüsselten Logomeren enthaltenen Informationen kann das Vorhandensein der Schlüsselsequenz selbst Auskunft darüber geben, ob ein gekennzeichnetes Produkt echt ist, oder nicht. Die Information über die Authentizität des gekennzeichneten Produktes kann dabei über einen Prozeß, bei dem der Schlüsselprimer als Hybridisierungs-sonde einer Fluoreszenz-Nachweisreaktion dient, verifiziert oder falsifiziert werden.

30

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher die Verwendung informationstragender Polymere, insbesondere der erfindungsgemäß erhältlichen informationstragenden Polymere, zur Verschlüsselung von Information.

35

Die oben beschriebenen Verfahren ermöglichen beispielsweise einen Test der Qualität von Oligonukleotiden:

40

Ein typisches Problem bei der Herstellung von Oligonukleotiden, z.B. für molekularbiologische Zwecke, ist die Qualitätskontrolle derselben, die aufgrund des Herstellungsprozesses und der schwankenden Qualität der verwendeten Chemikalien problematisch ist.

Das in den erfindungsgemäßen Schritten I - V beschriebene Verfahren kann benutzt werden, die Qualität von Oligonukleotiden zu testen. Dazu wird eine binäre Grammatik wie die für den Zufallszahlengenerator (s.o.) oder eine unäre Grammatik wie die zur Herstellung von Molekulargewichtsstandards (s.u.) zur Erzeugung beliebig langer Logomere verwendet. Bei ansonsten identischen Bedingungen ist dann die Länge der aus einer Symbolpolymerisation (Schritt V des erfindungsgemäßen Verfahrens) erhaltenen Logomere direkt proportional zur Qualität der verwendeten Oligonukleotide: Je länger die elektrophoretisch sichtbar gemachten Logomere sind, desto besser ist die Qualität der verwendeten Oligonukleotide. Ein Beispiel für die aus einer Symbolpolymerisation erhaltene Leiter unterschiedlich langer Logomere findet sich unter Abbildung 3.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher die Verwendung informationstragender Polymere, insbesondere der erfindungsgemäß erhältlichen informationstragenden Polymere, zur Qualitätskontrolle synthetisch hergestellter Oligonukleotide.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Herstellung von Logomeren als Marker und Signaturen

Aufgrund der Fähigkeit, prinzipiell beliebige Informationen darzustellen, lassen sich die hier beschriebenen Logomere als Marker verwenden, um Erzeugnisse, Produkte, Stoffe und Geräte zu kennzeichnen.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher die Verwendung informationstragender Polymere, insbesondere der erfindungsgemäß erhältlichen informationstragenden Polymere, als Marker oder Signaturen.

Die Logomere können als "binäres Etikett" verwendet werden, das einem Erzeugnis hinzugefügt wird und Informationen über das derart gekennzeichnete Erzeugnis enthält. Die Logomere können dabei z.B. Informationen über

- a) Hersteller
- b) Produkt (z.B. Seriennummer)
- c) Verwendungszweck
- d) Stoffklasse
- e) Gefahrenklasse
- f) Qualität
- g) Reinheit
- h) Produktions- und Verfallsdatum

enthalten. Es lassen sich damit nahezu beliebige Produkte und Erzeugnisse mit nahezu beliebigen Informationen versehen. Da die Logomere ungiftig und biologisch abbaubar sind,

können sie auch für hochsensible Produkte wie Nahrungsmittel, Medikamente und pharmazeutische Erzeugnisse eingesetzt werden.

Gekennzeichnete Erzeugnisse können z.B. sein:

- 5 a) chemische Erzeugnisse, z.B. Lacke, Farben, Öle, Schmier- und Treibstoffe, Lösungsmittel, Tinte usw.
- b) flüssige Materialien: Lösungen, Suspensionen, Emulsionen
- c) gentechnische Erzeugnisse
- d) Nahrungsmittel, z.B. zur Kennzeichnung gentechnisch veränderter Bestandteile oder dem Monitoring von Nahrungsmitteln während des Produktionsprozesses
- 10 e) Medikamente und pharmazeutische Erzeugnisse
- f) Papiererzeugnisse, Dokumente, Geld
- g) Geräte

Der Bedarf zur Kennzeichnung von Produkten besteht in den unterschiedlichsten Bereichen und aufgrund unterschiedlicher Anforderungen. Gründe für eine Kennzeichnung können z.B. sein: Qualitätssicherung beim Produktionsprozeß, Überwachung der Produktreinheit und Vermeidung von Kontaminationen, Schutz vor Fälschung und Produktpiraterie, Nachweis bestimmter Produktbestandteile, Produktkennzeichnung mit beliebigen Produktinformationen. Dieser Bedarf stellt sich auch in besonders sensiblen Bereichen, etwa bei der Kennzeichnung von Nahrungsmitteln, medizinischen und pharmazeutischen Erzeugnissen und gentechnisch hergestellten oder veränderten Erzeugnissen. Hier ist es wünschenswert, daß Informationen über das jeweilige Produkt direkt am Produkt verfügbar sind, insbesondere wenn Verwechselungen oder Fälschungen vermieden werden sollen und auch die Identifikation unterschiedlicher Bestandteile eines Erzeugnisses gewünscht ist.

25 Es gibt zahlreiche Beispiele für den Bedarf von Produktkennzeichnungen. Z.B. können Logomere als Seriennummern eingesetzt werden, die in Autolacke gemischt werden, wodurch der Fahrzeughalter im Versicherungsfall (z.B. Unfall oder Diebstahl) identifiziert werden kann. Durch Produktkennzeichnungen könnte die Qualität und Reinheit chemischer Erzeugnisse überwacht und Kontaminationen vermieden werden. Ein permanentes Problem ist die Fälschung von Dokumenten, Unterschriften und Geld, die hochentwickelte Kennzeichnungsverfahren nötig macht.

30 Vor allem vor dem Hintergrund von Tierkrankheiten und -seuchen (BSE, Schweinepest, Salmonellenvergiftungen usw.) stellt sich das Problem der Qualitätssicherung von Nahrungsmitteln und der Kennzeichnung von Endprodukten, um kontaminierte Produkte vom Verbrauch fernzuhalten. Ein weiteres Problem sind Nahrungsmittel, die gentechnisch hergestellte Bestandteile enthalten. Diese Bestandteile können ohne ein Kennzeichnungsverfahren im Endprodukt überhaupt nicht oder nur sehr schwer nachgewiesen werden. Das Problem stellt sich insbesondere dadurch, daß viele Nahrungsmittel Bestandteile unterschiedlicher Herkunft enthalten und die Produktionswege teilweise undurchschaubar sind.

40

Das Problem der Kennzeichnung stellt sich z.B. auch im Bereich von Blutprodukten, wo z.B. durch Pooling Kontaminationen auftreten können. Auch wäre es hier z.T. wünschenswert, Informationen über Identität, Herkunft, Produktions- und Verfallsdatum direkt mit dem Produkt verfügbar zu haben. Ähnliches gilt auch für Medikamente und pharmazeutische Produkte.

5

Bisher verwendet man zur Kennzeichnung und Markierung von Farben, Lacken, Treibstoffen, usw. verschiedene Kohlenstoffverbindungen, Polyaniline, Flüssigkristalle oder andere chemische Verbindungen. Diese Stoffe und Verbindungen haben unterschiedliche Schwächen: sie sind z.T. toxisch, nur schwer biologisch abbaubar, die verfügbare Informationskapazität ist stark eingeschränkt, sie wechselwirken chemisch mit bestimmten Stoffen oder sind nur teuer und aufwendig herzustellen.

10

Für den Einsatz zur Kennzeichnung von sensiblen und hochsensiblen Erzeugnissen wie Nahrungsmitteln oder Medikamenten sind die genannten Stoffe und Verbindungen aufgrund der genannten Nachteile überhaupt nicht geeignet.

15

Die Anforderungen, die sich für ein Kennzeichnungsverfahren von beliebigen Erzeugnissen stellen umfassen zumindest:

- Die Kennzeichnung sollte direkt im oder am Erzeugnis verfügbar sein;
- die Kennzeichnung sollte gut nachweisbar sein;
- die Kennzeichnung muß gesundheitlich unbedenklich sein.

20

Gegenüber den bisher verwendeten Verbindungen haben die hier beschriebenen Logomere folgende Vorzüge:

25

- Kodierung beliebiger Informationen,
- hohe Speicherkapazität,
- gesundheitlich unbedenklich,
- chemisch, biologisch, pharmakologisch und genetisch neutral,
- leicht biologisch abbaubar (Biomoleküle),
- keine Umweltbelastung,
- sehr gute Nachweisbarkeit,
- hohe Kompatibilität zu bestehenden Techniken der Informatik und Molekularbiologie (PCR),
- Informationen können verschlüsselt werden und eignen sich auch zu Authentifizierung und Verschlüsselung,
- Identifikation einzelner Bestandteile in Produktmischungen,
- billig in großen Mengen herzustellen (Klonierung, Vervielfältigung mit Bakterien in Fermentern).

30

35

Diese Eigenschaften machen Logomere für die genannten Aufgaben sehr viel besser geeignet als die bestehenden Techniken. Überdies erschließen sich Einsatzbereiche, die mit den bisherigen Techniken nicht möglich waren.

49

Aufgrund der Eigenschaften nukleinsäurebasierter Logomere können diese auch zur Kennzeichnung von besonders sensiblen Produkten wie Nahrungsmitteln und pharmazeutischen Erzeugnissen verwendet werden.

Die Gründe dafür liegen zum einen in der chemischen Beschaffenheit nukleinsäurebasierter

5 Logomere und zum anderen in der Kodierung von Informationen in Logomeren:

- 1) Als "Grundbausteine des Lebens" sind Nukleinsäuren Elementarbausteine aller bisher bekannten biologischen Lebensformen. Daher können sie aufgrund ihrer *chemischen* Beschaffenheit für keinen bekannten biologischen Organismus schädlich sein ("biochemische Kompatibilität"). Aufgrund der enthaltenen genetischen *Informationen* (des
10 enthaltenen Programmes) können Nukleinsäuren jedoch sehr wohl potentiell schädlich sein, wenn sie in einem Organismus abgelesen werden: Falls sie etwa für giftige Proteine kodieren oder (wie im Falle viralen Erbgutes) einen jeweiligen Zielorganismus "umprogrammieren" können. Die hier verwendeten Logomere können jedoch aus den im Folgenden genannten Gründen keine für einen Organismus schädlichen Informationen
15 enthalten.
- 2) Logomere enthalten keine genetischen, d.h. biologisch relevanten Informationen ("semantische Inkompatibilität"): Bezüglich der enthaltenen Informationen entsprechen Logomere Buchstaben, Zahlen oder Folgen von Buchstaben und Zahlen in Binärkodierung, die für uns oder einen Computer, jedoch nicht für den genetischen
20 Apparat eines Organismus lesbar sind. Für einen biologischen Organismus sind sie schlicht "Unsinnscodes".
- 3) Logomere, insbesondere die Algomere, aus denen sie aufgebaut sind, sind zu kurz um auch nur *zufällig* biologisch relevante Informationen zu enthalten.
- 4) Um jede Eventualität (der unwahrscheinliche Fall, daß die Binärinformationen der
25 Logomere auf irgendeine Weise von irgendeinem Organismus als genetisch relevante Information interpretiert wird) auszuschließen, werden die zur Kodierung von Symbolen verwendeten Sequenzen im Falle sensibler Anwendungsbereiche mit Stopcodons versehen, so daß sie niemals von einem biologischen Organismus abgelesen werden können.
- 5) Als elementare Bestandteile aller biologischen Organismen sind Nukleinsäuren in unserer
30 Umwelt allgegenwärtig. Sie werden in kürzester Zeit abgebaut.

Um als Marker eingesetzt zu werden, können Logomere nach den in den erfindungsgemäßen Schritten I - V beschriebenen Verfahren erzeugt und nach den erfindungsgemäßen oben
35 beschriebenen Verfahren vervielfältigt werden. Je nach Einsatzbereich können zur Vervielfältigung der Logomere verschiedene Verfahren und verschiedene Aufarbeitungsprozesse vorgenommen werden. Zur Kennzeichnung petrochemischer Erzeugnisse z.B. können Logomere durch Klonierung in Bakterien vervielfältigt werden. Um unnötige Kontaminationen zu vermeiden, können die erhaltenen Bakterien lysiert werden.
40 Eine spezielle Aufreinigung der Logomer-DNA ist im Allgemeinen nicht nötig.

- Zur Herstellung von Logomeren für sensible und hochsensible Produkte ist dagegen ein aufwendigeres Vervielfältigungs- und Aufreinigungsverfahren nötig. Ziel ist dabei der Erhalt möglichst reiner Logomere. Um biologische Zwischenschritte zu vermeiden, kann die Vervielfältigung - wie erfindungsgemäß beschrieben - mit PCR vorgenommen werden. Sollte
- 5 eine Vervielfältigung durch Klonierung vorgenommen werden, so ist die Aufreinigung der erhaltenen DNA aus Bakterien nötig. Sind die Logomere in bakteriellen Vektoren kloniert, so ist ein gezielter Abbau dieser Resistenzgene (z. B. durch Restriktionsenzyme) Bestandteil einer weiteren Aufarbeitung.
- 10 Logomere können zur Kennzeichnung eines Produktes unmittelbar mit diesem verbunden werden. Sie können z.B. in flüssige Stoffe gemischt werden oder auf feste Stoffe aufgebracht werden. Im Falle gentechnischer Erzeugnisse können die Logomere mithilfe von Rekombinations- und Klonierungstechniken kovalent mit dem zu kennzeichnenden Produkt verbunden werden.
- 15 Die Kennzeichnung von Stoffen und Produkten durch Logomere ermöglichen u.a. das Monitoring von Produktionsprozessen, die Qualitätskontrolle, die Identifikation einzelner Bestandteile und den quantitativen und qualitativen Nachweis von Kontaminationen.

- Ein Aspekt der vorliegenden Erfindung wird im folgenden am Beispiel der Herstellung
- 20 Computer-kompatibler Daten (Bytes/Multibytes) aus Nukleinsäuren und ihrer Verwendung zur Markierung unterschiedlicher Materialien erläutert, schränkt die Erfindung jedoch nicht auf dieses Beispiel ein:

1. Es wird eine Kodierung von Bytes und Multibytes in Form von Nukleinsäuren definiert.
- 25 2. Geeignete Sequenzen zur Repräsentation von Bytes und Multibytes werden konstruiert.
3. Eine Bibliothek (Library) von Byte-repräsentierenden Nukleinsäuren wird in vitro hergestellt.
4. Bytes werden zu Multibytes verknüpft.
5. Biochips werden zur Identifikation von Bytes und Multibytes hergestellt.
- 30 6. Unterschiedliche (organische und anorganische) Materialien und Gegenstände sowie einzelne Gene werden mit Bytes oder Multibytes gekennzeichnet (und ggfs. verschlüsselt).
7. Die Information gekennzeichneten Materialien, Gegenstände oder einzelner Gene wird ausgelesen (und ggfs. zu entschlüsselt).

- 35 Erläuterung zu 1:

Die Repräsentation von Daten mit Nukleinsäuren soll gewährleisten, daß die Daten Computer-kompatibel sind. Zu diesem Zweck wird als universellste Darstellung von Computerdaten die Repräsentation von Bytes und Multibytes gewählt.

Dazu wird eine Grammatik G_{C32} definiert, $G_{C32} = (\Sigma, V, R, S)$ mit Terminalalphabet $\Sigma := \{x_n, o_m, s_m, e_m\}$, Variablenalphabet $V := \{A, B, C\}$, Startsymbol S und Regelmenge $R := \{S \rightarrow oA, A \rightarrow s_mB, B \rightarrow Ce_m, C \rightarrow x_n\}$, wobei $n, m \in \mathbb{N}$, $n = \{0, 1, \dots, 255\}$ und $m \geq 0$. Die Grammatik beschreibt die Konstruktion von Molekülen der Form $osxe$, wobei x alle Bytewert zwischen 0 und 255 annehmen kann, s Bytepositionen beschreibt und o und e für das Verketteten von Einzelbytes zu Multibytes vorgesehen ist (siehe Abbildung 29 und Abbildung 30). Zur Vereinfachung des Herstellungsprozesses werden Restriktionssites als Variablen verwendet.

Außerdem müssen die verwendeten Nukleinsäuren logische, physikalische, chemische und biologische Eigenschaften haben. Insbesondere müssen/sollen sie einfach zu vervielfältigen, einfach zu lesen sein und auf definierte Weise mit biologischen Sequenzen interagieren.

Zu diesem Zweck werden datenrepräsentierende Nukleinsäuren so definiert, daß

- a) sie Bytes darstellen
 - b) auch komplexere Datenstrukturen (Multibytes) darstellen können
 - c) einzelne Bytes zu Multibytes verknüpft werden können
 - d) sie mit einem Biochip gelesen werden können
 - e) sie selbst zur Herstellung von 1-Byte und Multibyte Biochips dienen können
 - f) einzelne Bytes und Multibytes mit PCR vervielfältigt werden können
 - g) klonierbar sind
 - h) mit anderen Nukleinsäuren so verknüpft werden können, daß sie diese markieren
 - i) möglichst wenige Restriktionssites tragen
- a) Damit die datenrepräsentierenden Nukleinsäuren Bytes darstellen können, müssen genau 256 eindeutige Sequenzen alle Werte eines Bytes repräsentieren. Dazu enthalten die molekularen Bytes eine eindeutige Sequenz x , von der es 256 verschiedene Allele gibt. Diese Allele werden als x_0 bis x_{255} bezeichnet und enumerieren genau alle Bytewerte ($x_0 = 0, x_1 = 1, \dots, x_{255} = 255$) (siehe Abbildung 25, Abbildung 26, Abbildung 27, Abbildung 28 und Sequenzprotokoll).
- b) Um komplexere Datenstrukturen (Multibytes) wie z.B. 32-Bit Zahlen und Strings darstellen zu können, werden die Bytemoleküle so konstruiert, daß sie mit einer Positionsinformation versehen werde. Diese ist in Form der Sequenz s enthalten (siehe Abbildung 25, Abbildung 26, Abbildung 27, Abbildung 28 und Sequenzprotokoll). Jede Position ist durch eine eindeutige Sequenz s_i dargestellt. Mit den verschiedenen Allelen von s werden dann alle Bytepositionen enumeriert (s_0 = Position 0, s_1 = Position 1, ... usw.).
- c) Obwohl es im Prinzip zur Darstellung von Multibytes ausreicht, die darzustellenden Daten auf unabhängige Bytes mit den entsprechenden Positionsinformationen zu verteilen (als mehrsträngige Multibytes: z.B. können, um einen Lack mit einer 4 Byte langen Seriennummer zu versehen, 4 einzelne Bytes die jeweils genau eine Positionsinformation von s_0 bis s_3 besitzen unabhängig voneinander in den Lack gemischt werden), ist es manchmal zweckmäßig, Bytes direkt zu einsträngigen Multibytes zu verknüpfen. Dazu tragen die Bytes die Sequenzen o und e (siehe Abbildung 25 bis Abbildung 28 und Sequenzprotokoll), mit denen sie sich durch Ligation oder overlap assembly und PCR zu

einsträngigen Multibytes verknüpfen lassen (siehe Abbildung 29 und Abbildung 30). Diese einstängigen Multibytes können dann ihrerseits als Marker verwendet werden. Im Falle des parallel overlap assembly werden die einzelnen Bytes mit einem geeigneten Restriktionsenzym (hier: EcoRV siehe Abbildung 28) ausgeschnitten.

- 5 d) Datenrepräsentierende Nukleinsäuren können z.B. mithilfe von Sequenzierern gelesen werden. Damit sie jedoch möglichst einfach und schnell gelesen werden können, ist insbesondere vorgesehen, sie mithilfe von Biochips zu lesen. Dazu werden Biochips hergestellt, die alle 256 verschiedenen x-Sequenzen einzelsträngig (und damit alle Bytewerte) aufsteigend geordnet enthalten (siehe Abbildung 32 und Abbildung 33).
10 Solchermaßen hergestellte Biochips sind ebenfalls Gegenstand der vorliegenden Erfindung. Das Auslesen eines bestimmten Bytewertes x_i erfolgt dann durch Denaturieren des entsprechenden Datenmoleküls zu Einzelsträngen und Hybridisieren mit dem Chip. Dabei werden die hybridisierten Sequenzen analog der typischen Funktionsweise von Biochips markiert, z.B. durch Fluoreszenzmarkierung. Nach dem Hybridisieren von
15 Datenmolekül mit einem 1-Byte Chip enthält man dadurch genau eine markierte Position auf dem Biochip, die damit genau den Bytewert bezeichnet (siehe Abbildung 36). Z.B. hybridisiert das Datenmolekül, das die Sequenz x_{192} trägt, genau an der Position 192 im Chip, und kann anhand der Position innerhalb des Chips identifiziert werden.
- 20 e) Da das Prinzip des Auslesens von Datenmolekülen mithilfe des Biochips auf der Hybridisierung komplementärer Sequenzen beruht, werden die Bytes so konstruiert, daß sie gleichzeitig zur Herstellung der entsprechenden Biochips dienen können. Dazu tragen die Bytemoleküle enzymatische Restriktionssites, so daß aus einem vollständigen Bytemolekül entweder die x enthaltende Subsequenz oder die s und x enthaltende Subsequenz ausgeschnitten werden können. Diese Subsequenzen werden dann
25 denaturiert und auf einen Chipträger gespottet. Um einen 1-Byte Chip herzustellen werden alle 256 x-Sequenzen aufsteigend geordnet auf den Chip aufgetragen. Dadurch ist es möglich, eine Sequenz mit vorher unbekanntem x anhand ihrer Hybridisierungsposition zu identifizieren. Verwendet man nur die x Subsequenz für die Herstellung des 1-Byte Chips, so enthält man als X-Chips bezeichnete Chips (siehe Abbildung 32), verwendet
30 man die s und x enthaltende Subsequenz, so erhält man die als SX-Chips bezeichnete Chips (siehe Abbildung 33). Sowohl X- als auch SX-Chips können auch zu Multibyte Chips kombiniert werden. Eine weitere Voraussetzung für das Herstellen von Biochips ist die Verfügbarkeit ausreichender Mengen von Sequenzen. Dazu kann man die Bytemoleküle entweder direkt mit Oligosynthesizern herstellen, sie klonieren oder aus
35 wenigen Template-Molekülen mit PCR vervielfältigen.
- f) Um gezielt einzelne Bytes (z.B. aus einsträngigen Multibytes) zu vervielfältigen, können die als o, s, und e bezeichneten Subsequenzen als priming sites dienen (siehe Abbildung 25, Abbildung 26, Abbildung 27, Abbildung 28).
- 40 g) Um klonierbar zu sein tragen die Bytemoleküle sticky ends die zu Restriktionssites kompatibel sind (z.B. XhoI und XbaI, siehe Abbildung 28).

- h) Damit Bytes und Multibytes leichter mit beliebigen Nukleinsäuren verknüpft werden können, können entweder durch Ligation oder durch overlap assembly und PCR weitere Terminatoren (Adaptoren) an Bytes und Multibytes angehängt werden (siehe Abbildung 29 und Abbildung 30). Diese Adaptoren tragen dazu Restriktionssites oder Rekombinationssites, so daß sie entweder durch Restriktion und Ligation oder durch
- 5 Rekombination mit anderen Nukleinsäuren verknüpft werden können. Dies ist insbesondere zur Kennzeichnung von Nukleinsäuren, insbesondere Genen vorteilhaft.
- i) Im Zusammenhang mit der Verknüpfung von Bytes und Multibytes mit anderen Nukleinsäuren ist es vorteilhaft, daß insbesondere Multibytes nur wenige Restriktionssites tragen, damit der Verwender von Nukleinsäuren die mit Bytes und Multibytes versehen sind, möglichst wie gewohnt damit arbeiten kann und möglichst auf keine Restriktionssites verzichten muß, die ansonsten die Bytes zerstören. Daher sind die Bytes so konstruiert, daß nach der Herstellung von einsträngigen Multibytes von allen Restriktionssites, die auf Sequenzpalindromen der Länge 6 beruhen nur noch die Restriktionssites Afill, NheI und
- 10 NcoI in den Multibytes enthalten sind und daher nicht ohne weiteres z.B. für mit Multibytes markierte Nukleinsäuren verwendet werden können. Ggfs. lassen sich sogar die noch verbleibenden Restriktionssites aus den Bytes und Multibytes entfernen ohne die Byte und Positionsinformationen zu verlieren, indem man Multibytes mit PCR und mutierten Primern so vervielfältigt, daß die Restriktionssites blockiert werden.

20

Erläuterung zu 2:

Die zur Implementierung benötigten Sequenzen werden mithilfe des bereits beschriebenen NFR-Verfahrens synthetisiert. Die Sequenzen sind so konstruiert, daß alle Sequenzen untereinander möglichst unähnlich sind. Die x-Subsequenzen

25 haben einen GC-Anteil von 50%, die o-, s- und e-Sequenzen haben einen GC-Anteil von 66%, Multibytes enthalten 3 Sequenzpalindrome der Länge 6, die jeweils einmal von Afill, NheI und NcoI erkannt werden.

Erläuterung zu 3: Die Bytemoleküle können direkt mit Hilfe eines Oligosynthesizers

30 hergestellt werden. Um größere Mengen von Bytemolekülen herzustellen ist es jedoch oft zweckmäßig, eine Klonbibliothek anzulegen, die alle Bytemoleküle enthält.

Erläuterung zu 4: Bytes können aufgrund der enthaltenen Bytepositionsinformation auch ohne direkt miteinander verknüpft zu werden Multibytes darstellen (logische Verknüpfung zu mehrsträngigen Multibytes). Es ist jedoch für einige Anwendungen,

35 z.B. die Kennzeichnung von Nukleinsäuresträngen zweckmäßig, die einzelnen Bytes auch physikalisch zu einem Multibyte zu verknüpfen (einsträngiges Multibyte, siehe Abbildung 30 und Abbildung 31). Dazu werden die einzelnen Bytes entweder durch Ligation oder durch overlap assembly und PCR miteinander verknüpft.

- Erläuterung zu 5: Biochips zum Lesen von Bytes und Multibytes werden mit Untereinheiten der Moleküle hergestellt, die als Bytes fungieren. Grundsätzlich gibt es zumindest 2 Architekturen: den X-Chip (siehe Abbildung 32), der mit x-Subsequenzen (siehe Abbildung 25 bis Abbildung 28) und den SX-Chip (siehe Abbildung 33), der mit sx-Subsequenzen (siehe Abbildung 25 bis Abbildung 28) hergestellt wird. Beide Chips sind 1-Byte Chips und können zu Multibyte Chips kombiniert werden. Die prinzipielle Funktionsweise ist ähnlich: Das Auslesen der Information von Datenmolekülen erfolgt durch Hybridisieren mit dem Chip. Anhand der Hybridisierungsposition des Datenmoleküls läßt sich sein Bytewert ablesen.
- 5
- 10
- 15
- 20
- 25
- Dadurch wird es auch möglich, Multibyte Chips als Datenspeicher (ROM und RAM) zu verwenden: Das Hybridisieren des Chips mit einem spezifischen Datenmolekül entspricht dem Zuweisen eines Wertes an eine Speicherzelle. Durch Denaturieren (z.B. über Temperaturerhöhung oder pH-Wert Änderung) wird die Speicherzelle wieder gelöscht. Darüber hinaus können spezifische Sequenzen die mit dem Chip hybridisiert werden auch mit optisch aktiven Molekülen unterschiedlicher Farbe versehen werden, so daß sich auch gezielt Farbmuster erzeugen lassen, so daß sich Multibyte Chips auch als Display verwenden lassen. Kombiniert man Sequenzen unterschiedlicher Schmelztemperatur mit optisch aktiven Molekülen verschiedener Farbe, so kann auch eine Farbanzeige im Sinne eines Thermometers erfolgen.
- Der Unterschied zwischen X-Chip und SX-Chip besteht darin, daß der SX-Chip zusätzlich zum Bytewert auch die Byteposition lesen (und speichern) kann. Vorteil des X-Chip ist, daß man sehr große Multibyte-Arrays aus identischen X-Chips herstellen kann. Vorteil des SX-Chips ist, daß sich damit einsträngige Multibyte-Moleküle lesen lassen, ohne daß man die einzelnen Bytes getrennt voneinander hybridisieren müßte.

- Erläuterung zu 6: Über die Adaptoren L und R (siehe Abbildung 29 und Abbildung 30) lassen sich Sequenzen mit Bytes und Multibytes verbinden, die dazu dienen, diese mit Nukleinsäuren zu verbinden. Dazu enthalten die Adaptoren z.B. geeignete Restriktionssites oder Rekombinationssites. Z.B. läßt sich das ein Multibyte mithilfe von geeigneten Restriktionssites in einem Plasmid klonieren, wodurch dieses Plasmid eine Kennzeichnung erhält. Diese kann z.B. in einer 32-Bit Seriennummer bestehen (siehe Abbildung 31). Ein anderes Anwendungsbeispiel ist das Anbringen von Rekombinationssites an ein Datenmolekül, wodurch dieses z.B. in nicht-kodierende Regionen (z.B. Introns) von Genen eingebracht werden kann, wodurch diese gekennzeichnet werden.
- 30
- 35

Erläuterung zu 7: Die Information von Materialien, Gegenständen und Molekülen, die mit Datenmolekülen gekennzeichnet wurden lassen sich dadurch auslesen, daß man sie aus dem gekennzeichneten Material extrahiert, ggfs. aufreinigt, ggfs. mit PCR
5 amplifiziert und dann entweder sequenziert oder durch Hybridisieren mit einem Byte- oder Multibyte-Chip wie oben identifiziert.

Die oben beschriebenen Verfahren ermöglichen somit die eindeutige Kennzeichnung von organischen und anorganischen Materialien und Gegenständen sowie von
10 Nukleinsäure-Konstrukten und Genen.

Die oben beschriebenen Verfahren ermöglichen außerdem die Herstellung von Datenspeichern und optischen Displays: Zum Zwecke der Speicherung von Computerdaten werden Multibyte-Arrays verwendet. Das Schreiben eines einzelnen
15 Byte erfolgt durch das Hybridisieren eines 1-Byte Chips mit einer (ggfs. optisch markierten) Nukleinsäure, die genau einmal eine definierte x-Sequenz enthält, die dem zu schreibenden Wert entspricht (z.B. x_{192}). Das Auslesen der Daten erfolgt analog dem Auslesen biologischer DNA-Arrays, z.B. durch Scannen. Das Löschen von Daten erfolgt durch Denaturieren und ggfs. nachträgliches Waschen.

20

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Kennzeichnung einzelner Moleküle, Nukleinsäuren und gentechnisch hergestellter oder veränderter Erzeugnisse:

25

Die hier beschriebenen Logomere können zur Kennzeichnung gentechnisch hergestellter oder veränderter Erzeugnisse verwendet werden.

Bisher werden insbesondere zur Identifikation gentechnisch hergestellter oder veränderter Erzeugnisse in Nahrungsmitteln "klassische" molekularbiologische Methoden wie PCR,
30 Restriktionsverdau, Southernblot-Hybridisierung und Sequenzierung verwendet. Mit diesen Methoden ist jedoch eine Charakterisierung gentechnischer Erzeugnisse und Bestandteile noch sehr aufwendig und erfordert für jedes zu charakterisierende Erzeugnis eigene Methoden und Verfahren.

Insgesamt gibt es bisher kein universelles Kennzeichnungsverfahren für gentechnisch
35 hergestellte Erzeugnisse. Ein solches Kennzeichnungsverfahren muß mehrere Kriterien erfüllen:

- die Kennzeichnung muß gesundheitlich absolut unbedenklich sein,
- die Kennzeichnung sollte untrennbar mit dem gekennzeichneten Produkt verbunden sein,
- die Kennzeichnung sollte fälschungssicher sein,

- die Kennzeichnung sollte genügend Informationen speichern können,
- die Kennzeichnung sollte in geringen Mengen nachweisbar und lesbar sein.

Der Einsatz nukleinsäurebasierter Logomere zur Kennzeichnung gentechnischer Erzeugnisse erfüllt diese Kriterien und hat darüber hinaus folgende Vorzüge:

- 5 - Logomere können nahezu beliebige Informationen speichern;
- einzelne Bestandteile können schnell und hochempfindlich nachgewiesen werden;
- zum Auslesen der Informationen kann eine bis ins Detail standardisierte Methode eingesetzt werden (Ausleseverfahren nach dem erfindungsgemäßen Verfahrens, dabei können standardisierte Primer verwendet werden);
- 10 - die Sequenz der gesuchten Gene kann vollkommen unbekannt sein;
- die Sequenz der nachgewiesenen Gene kann geheim sein, ohne daß die Identifizierung oder Authentifizierung eingeschränkt ist;
- zusätzliche Sicherheitsmechanismen können zusätzlich durch Verschlüsselung implementiert werden.

15

Im Einzelnen geht man zur Kennzeichnung von gentechnisch erzeugten oder veränderten Erzeugnissen beispielsweise folgendermaßen vor:

Zur Herstellung von Markern wird eine beliebige geeignete Grammatik, z.B. wie die für den Zufallszahlengenerator, das 1-Byte Alphabet oder für Zeichenketten bereits beschriebene, verwendet. Aus den mit der Grammatik erzeugten Zeichen werden die zur Darstellung von Markern benötigten Zeichen entnommen und dem zu kennzeichnenden Erzeugnis hinzugefügt. Dazu können die als Logomere hergestellten Zeichen auf folgende Weise in das zu kennzeichnende Erzeugnis eingebracht werden:

20

25

30

35

40

- a) durch Beimischung; hierbei werden Logomere, wie in den erfindungsgemäßen Verfahren beschrieben, hergestellt, isoliert und vervielfältigt.
- b) durch Klonierung, wie im erfindungsgemäßen Verfahren beschrieben; Abbildung 6 zeigt z.B. die Kennzeichnung bakterieller Klone durch Logomere, die durch Verwendung der bereits beschriebenen Grammatik für den Zufallszahlengenerator hergestellt wurden.
- c) durch Restriktion und Ligation; hierbei werden Logomere erfindungsgemäß hergestellt, isoliert und vervielfältigt. Die solcherart gewonnenen Logomere werden jetzt durch Restriktion und Ligation mit der zu kennzeichnenden Nukleinsäure verbunden. Dazu können die oben beschriebenen Terminatoren (siehe Abbildung 2) und Adaptoren (siehe Abbildung 30) gewünschte Restriktionssites tragen.
- d) durch den Einsatz rekombinativer Techniken; hierbei werden Logomere erfindungsgemäß hergestellt, isoliert und vervielfältigt. Die solcherart gewonnenen Logomere werden jetzt durch den Einsatz rekombinativer Techniken in das zu kennzeichnende Erzeugnis eingebracht. Dazu kann z. B. die Methoden des Gene-Targeting durch homologe Rekombination [Capecchi M.R., Altering the genome by homologous recombination. *Science*, **244**(4910), 1288-1292, (1989)] oder z.B. das Cre-loxP System [Kilby, N. J., Snaith, M. R. & Murray, J. A., Site-specific recombinases: tools for genome engineering, *Trends Genet.*, **9**, 413-421, (1993)] verwendet werden. Z.B. können erfindungsgemäß

erhaltene Logomere mit Hilfe dieser Techniken vor oder hinter eine bestimmte Nukleinsäuresequenz (Gen) gesetzt werden, die sie bezeichnen sollen. In diesem Falle wird ein Logomer gewissermaßen als "Etikett" an ein gentechnisches Produkt "angeheftet" und kann Auskunft über Hersteller, Produktgruppe, Gefahrenklasse, Verfallsdatum o.ä. geben (siehe Abbildung 15).

Das Verfahren eignet sich insbesondere zur Kennzeichnung einzelner Moleküle, insbesondere einzelner Nukleinsäuren, besonders bevorzugt einzelner Gene.

Das Verfahren eignet sich z.B. für alle Organismen (Mikroorganismen, Pflanzen, Tiere). Z.B. lassen sich Klone von Mikroorganismen oder Pflanzen kennzeichnen. Das Verfahren eignet sich auch für die Kennzeichnung von Lebensmitteln und für die Kennzeichnung von Rindern zur Bekämpfung der Rinderseuche BSE.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Kennzeichnung von Nahrungsmitteln

Aufgrund von Krankheiten, Seuchen oder Kontaminationen können Nahrungsmittel gesundheitsschädlich oder gefährlich sein. Beispiele sind Krankheiten wie BSE, Schweinepest, Salmonellenvergiftungen. Es kann wünschenswert sein, Produkt und Hersteller identifizieren zu können, um notfalls bestimmte Produkte und Produktreihen umgehend aus dem Handel zu nehmen, untersuchen zu können und geeignete Gegenmaßnahmen zu ergreifen.

Für hochsensible Erzeugnisse wie Nahrungsmittel gelten dabei die Kriterien für die Kennzeichnung gentechnischer Erzeugnisse insbesondere (siehe: *Die oben beschriebenen Verfahren ermöglichen beispielsweise die Kennzeichnung einzelner Moleküle, Nukleinsäuren und gentechnisch hergestellter oder veränderter Erzeugnisse*).

Insgesamt gibt es auch für gentechnisch hergestellte Nahrungsmittel bisher kein universelles Kennzeichnungsverfahren. Bisher ist die Identifikation gentechnisch hergestellter oder veränderter Bestandteile in Nahrungsmitteln daher noch schwierig und aufwendig. Es werden dazu "klassische" molekularbiologische Methoden wie PCR, Restriktionsverdau, Southernblot-Hybridisierung und Sequenzierung verwendet, um diese Erzeugnisse zu charakterisieren.

Aufgrund der Rinderseuche BSE wurde speziell für die Kennzeichnung von Rindern und deren Produkten (Rindfleisch und Milch) ein immunologisches Kennzeichnungsverfahren vorgeschlagen, das auf der Immunisierung der zu markierenden Tiere mit spezifischen Proteinen basiert ("Ohrmarke im Blut soll BSE Rinder aufspüren", Süddeutsche Zeitung Nr. 283, 08.12.1998, Seite v2/9). Durch die dadurch verursachte Produktion spezifischer Antikörper im jeweils gekennzeichneten Zieltier, die mutmaßlich in allen Geweben nachzuweisen ist, kann die Kennzeichnung über einen Immun-Assay (ELISA) wieder gelesen werden. Ein solches Verfahren ist jedoch prinzipiell auf höhere Wirbeltiere beschränkt. Außerdem ist die für das Verfahren verfügbare Informationsmenge sehr begrenzt, die

informationstragenden Proteine sind nicht hitzebeständig und das Verfahren erfordert vergleichsweise hohe Mengen Substanz für Immunisierung und Nachweisreaktion.

- Auch im Fall von Nahrungsmitteln können Logomere auf Nukleinsäurebasis, wie bereits unter
- 5 *Die oben beschriebenen Verfahren ermöglichen beispielsweise die Kennzeichnung einzelner Moleküle, Nukleinsäuren und gentechnisch hergestellter oder veränderter Erzeugnisse* beschrieben, zur Kennzeichnung eingesetzt werden. Logomere können dabei Informationen über Hersteller, Verfallsdatum, Seriennummer u.a. enthalten. Das Verfahren zur Kennzeichnung von Nahrungsmitteln ist dabei dasselbe, wie bereits unter *Die oben*
- 10 *beschriebenen Verfahren ermöglichen beispielsweise die Kennzeichnung einzelner Moleküle, Nukleinsäuren und gentechnisch hergestellter oder veränderter Erzeugnisse* beschrieben. Mithilfe der dort aufgeführten gentechnischen Rekombinations- und Klonierungstechniken können einzelne Organismen gekennzeichnet werden, sodaß ein einzelner Organismus sowie alle seine Nachfahren identifizierbar sind.
- 15 Das Kennzeichnungsverfahren durch Logomere ermöglicht dadurch:
- ein Monitoring des Zielproduktes über den gesamten Herstellungs- und Verarbeitungsprozesses bis zum Endverbraucher,
 - eine einheitliche, schnelle und einfache Identifikation einzelner Bestandteile.
- 20 Als Beimischung eignen sich Logomere auf Nukleinsäurebasis zur Kennzeichnung von Getränken.

- Die oben beschriebenen Verfahren ermöglichen beispielsweise die Kennzeichnung
- 25 medizinischer und pharmazeutischer Produkte

- Aufgrund der Nichttoxizität eignen sich Logomere auf Nukleinsäurebasis zur Kennzeichnung sensibler und hochsensibler Erzeugnisse. Neben Nahrungsmitteln sind dies auch medizinische und pharmazeutische Produkte. Um Verwechselungen, Kontaminationen und
- 30 Fälschungen ausschließen zu können, können Logomere z.B. medizinischen und pharmazeutischen Produkten beigemischt werden.

- Der Vorteil der Kennzeichnung durch Logomere ist dabei, daß die Kennzeichnung unmittelbar mit dem gekennzeichneten Erzeugnis verbunden ist und daß sich Erzeugnisse z.B. individuell kennzeichnen lassen. Auf diese Weise ist ein Monitoring des Herstellungsprozesses möglich,
- 35 Erzeugnisse lassen sich auch nach mehrmaligem Umfüllen eindeutig identifizieren und Kontaminationen sowie das Poolen von Proben können nachgewiesen werden. Mithilfe von Logomeren läßt sich z.B. im Falle von Blutkonserven Herstellungs- und Verfallsdatum, Blutgruppe und Hersteller kennzeichnen.

Die oben beschriebenen Verfahren ermöglichen beispielsweise Authentifizierung und Fälschungssicherung

Logomere können zur Authentifizierung von Erzeugnissen, Gegenständen und Geräten verwendet werden. Dazu werden Logomere vorzugsweise erfindungsgemäß erzeugt und vervielfältigt.

Werden solcherart erhaltene Logomere auf Erzeugnisse, Gegenstände und Geräte an-/aufgebracht oder zugemischt, so können die in Logomeren enthaltenen erfindungsgemäß jederzeit wieder ausgelesen werden. Sollten zur Erzeugung der Logomere Verschlüsselungstechniken verwendet werden, wie bereits beschrieben, so erfordert das Auslesen autorisierten Zugriff.

Beispielsweise können erfindungsgemäß erhaltene Logomere in wässriger Lösung (vorzugsweise zu 10^3 bis 10^{18} Molekülen/ μ l) unmittelbar zur Kennzeichnung auf Dokumente aufgebracht werden. Als solches können sie als Echtheitszertifikat des solcherart markierten Dokumentes dienen. Zum Auslesen der als Kennzeichner dienenden Logomere wird einfach ein kleines Schnipsel des Papiers (1mm^2 ist ausreichend) als Template einer erfindungsgemäßen PCR-Auslesereaktion verwendet. Ein Beispiel findet sich in Abbildung 16. In dem zugrundeliegenden Ausleseexperiment wurde zum Test des Prinzips ein Logomer in wässriger Lösung in ca. 10^9 Molekülen/ μ l auf 3M PostIt Papier ca. 1 Stunde getrocknet und 1mm^2 des Post-Its als Template einer Auslese-PCR verwendet. Es können Dokumente beliebiger Papierqualität verwendet werden, z.B. 80g/m^2 Standardkopierpapier.

Dasselbe Verfahren eignet sich auch zur Kennzeichnung von Geld oder Geldscheinen, wobei hier Vorteile darin bestehen,

- auch bereits gedruckte Scheine markieren zu können,
- Seriennummern vergeben zu können,
- Verschlüsselung einzusetzen.

Um Dokumente möglichst in Echtzeit auf Authentizität prüfen zu können, kann einer der zum Auslesen in der PCR benutzten Primer auch als Hybridisierungs-sonde für eine nachfolgende Färbe-Nachweisreaktion (z.B. mit einem interkalierenden Farbstoff, z.B. Ethidiumbromid) verwendet werden.

Ein weiteres Beispiel ist die Authentifizierung flüssiger Lösungen, Suspensionen und Emulsionen mit Logomeren. So kann etwa Tinte mit Logomeren individuell markiert werden, so daß Unterschriften damit auf Echtheit überprüft werden können.

Ein weiteres Beispiel ist die Kennzeichnung von Automobilen durch das Versetzen des Autolackes mit Logomeren. Dabei werden Logomere dem Autolack (oder anderen Bestandteilen des KFZ) als Beimischung zugesetzt. Die zugesetzten Logomere können dabei Informationen über Seriennummer, Hersteller, Fahrzeugtyp o.a. Angaben enthalten. Aufgrund dieser Informationen kann im Falle von Unfall, Diebstahl oder illegaler Entsorgung ein Fahrzeug identifiziert werden. Ein Vorteil der hier beschriebenen Methode liegt darin, daß schon Lackspuren zur Identifikation des Fahrzeuges ausreichen, was insbesondere bei

Unfällen von Interesse sein kann. Ein weiterer Vorteil ist, daß die Identität eines Fahrzeuges nur sehr aufwendig zu fälschen ist. Dazu müßten erstens alle gekennzeichneten Bestandteile entfernt, zweitens eine gefälschte Identität eingerichtet werden. Dieses Unterfangen ist alleine durch den mechanischen Aufwand sehr schwierig, außerdem läßt sich durch den Einsatz

5 kryptografischer Methoden das Nachmachen und Nachahmen von Kennungen beliebig erschweren. Zum Auslesen der als Logomere enthaltenen Informationen, werden die Logomere aus dem Lack isoliert und können dann wie in der Beschreibung zum Auslesen der in Logomeren enthaltenen Information, vorzugsweise durch PCR, ausgelesen werden. Dazu wird eine Probe des Lackes entnommen, mit einem geeigneten Lösungsmittel (Terpentin o.ä.)

10 gelöst und mit gleichem Volumen Wasser versetzt. Durch nachfolgende Zentrifugation werden hydrophile und hydrophobe Bestandteile getrennt. Daraufhin nimmt man die wässrige Phase auf (worin sich die hydrophilen nukleinsäurebasierten Logomere befinden) und fällt die enthaltenen Logomere nach den üblichen Methoden (z.B. wie in [J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning, A Laboratory Manual*, (1989)] beschrieben). Nach der Fällung

15 können die erhaltenen Logomere nach den erfindungsgemäßen Verfahren, z.B. durch Auslese-PCR, ausgelesen werden.

Daher sind weitere Gegenstände der vorliegenden Erfindung die Verwendungen informationstragender Polymere, insbesondere der erfindungsgemäß erhältlichen

20 informationstragenden Polymere, zum Zwecke der Qualitätssicherung, der Fälschungssicherung, der Kennzeichnung von Lebensmitteln, der Kennzeichnung gentechnischer, chemischer, medizinischer und pharmazeutischer Erzeugnisse, der Kennzeichnung von Organismen, der Kennzeichnung von Dokumenten, der Kennzeichnung von Geld, der Kennzeichnung von Gegenständen und Maschinen, der Kennzeichnung von

25 Lösungen, Suspensionen und Emulsionen sowie der Authentifizierung von Personen und Gegenständen.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Herstellung von

30 Molekulargewichtsstandards

Zu den typischen Verfahren der molekularbiologischen Analytik gehört die elektrophoretische Längenauftrennung von DNA- und RNA-Fragmenten. Um die Länge solcher Fragmenten bei der Elektrophorese zu bestimmen, verwendet man üblicherweise einen

35 Molekulargewichtsstandard ("Leiter"), der Fragmente bekannter Länge enthält und zum Vergleich mit den Fragmenten der zu analysierenden Proben dient.

Bisher gibt es Molekulargewichtsstandards, die entweder unregelmäßige Leitern enthalten (z.B. Boehringer V, Boehringer Mannheim, Katalog Nr.: 821 705) oder regelmäßige Leitern mit regelmäßigen Fragment-Längenabständen (z.B. 50bp Leiter Gibco BRL, Life Technologies,

40 Katalog Nr. 10416-014). Keine der bisherigen Leitern hat dabei kürzere Längenabstände als 10bp (z.B. 10bp Leiter Gibco BRL, Life Technologies, Katalog Nr. 10821-015).

Mit den hier beschriebenen Verfahren können gezielt DNA Fragmente verschiedener Länge und definierter Längenabstände hergestellt werden. Diese Fragmente können als Molekulargewichtsstandards genutzt werden, wobei das Verfahren erlaubt, prinzipiell beliebige Längen und beliebige Längenabstände zu realisieren. Insbesondere sind Längenabstände bis

5 hinab zu 1bp Abstand, also Molekulargewichtsstandards höchster Auflösung, möglich.

Das Verfahren beruht darauf, Logomere als Template einer Auslese-PCR zu verwenden, deren Resultat ein Gemisch von DNA Fragmenten definierter Länge und definierter Längenabstände ist. Es basiert auf den bereits beschriebenen erfindungsgemäßen Verfahren zur Erzeugung und dem Auslesen von Logomeren.

10 DNA Fragmente verschiedener Länge in sehr kurzen, regelmäßigen Abständen können erhalten werden, wenn ein unäres Polymer als Template einer Auslese-PCR mit verschachtelten Primern fungiert. Der 5'-Primer primt in der Start-Sequenz des Polymers. Als gegenläufige 3' Primer fungiert eine Gruppe von Primern, die verschachtelt in dem Elongator ansetzt (in Abbildung 17 dargestellt als 3 Ebenen gegeneinander versetzter Primer). Da das

15 Polymer aus Wiederholungen nur eines Elongator-Moleküls besteht, erhält man für n Wiederholungen und m verschachtelte 3' Primer n*m Banden. Das Verfahren funktioniert auch spiegelbildlich mit einem im Ende-Terminator primenden Primer und dementsprechenden gegenläufigen 5' Primern. Die erhaltenen Banden können als Molekulargewichtsstandards z.B. in Elektrophoresen eingesetzt werden.

20

Für hochauflösende Molekulargewichtsstandards ist die Verwendung unärer Logomere und verschachtelter Primer am zweckmäßigsten. Jedoch ist das Verfahren nicht darauf beschränkt, da im Prinzip alle im erfindungsgemäßen Schritt I beschriebenen Grammatiken verwendet werden können.

25 Unäre Logomere beliebiger Länge können mit einer Grammatik $G = (\Sigma, V, R, S)$ mit Terminalalphabet $\Sigma := \{0, s, e\}$, Variablenmenge $V := \{A\}$, Startsymbol S und Regelmenge $R :=$

```
{
    S:=sA
    A->aA
    A->e
}
```

30

hergestellt werden. Sollte es nötig sein, unäre Logomere definierter Länge herzustellen, muß dementsprechend die Variablenmenge und damit auch die Zahl der Elongatoren größer

35 gewählt werden (wie z.B. in der Grammatik zur Implementierung eines 1-Byte Alphabets bereits beschrieben). Je nach gewünschter Länge der zu erzeugenden DNA Fragmente kann außerdem die Länge der Logomere variiert werden.

Die so erzeugten Logomere können nach den erfindungsgemäßen Verfahren vereinzelt und vervielfältigt werden.

40 Das Erzeugen von DNA-Fragmenten definierter Länge und definierter Längenabstände erfolgt vorzugsweise mit PCR analog dem erfindungsgemäßen Verfahren zum Auslesen von

Logomeren. Für hochauflösende Molekulargewichtsstandards wird pro Elongator mindestens ein Primer, meistens sogar eine Anzahl mehrerer Primer verwendet, die gegeneinander verschoben sind. Die Anzahl der Primer pro Elongator ist abhängig von den gewünschten Längenabständen. Ist etwa gewünscht, Algomere der Länge 30bp zu verwenden und sollen
5 die Längen der erzeugten DNA Fragmente jeweils 10bp Unterschied haben, so sind pro Elongator drei Primer, die jeweils genau 10bp gegeneinander verschoben sind, eingesetzt werden.

Für die Synthese von Sequenzen zur Herstellung von Algomeren und Logomeren müssen die Anforderungen erfüllt werden, die bereits unter Schritt II des erfindungsgemäßen Verfahrens
10 beschrieben wurden. Insbesondere gewährleistet das verwendete NFR-Verfahren dadurch, daß die jeweils als Template dienenden Teilsequenzen einen gleichen GC-Anteil besitzen können, den Einsatz verschachtelter Primer und die Ausgewogenheit des AT/GC Verhältnisses der erzeugten DNA Fragmente. Letzteres ist insbesondere angesichts dessen
15 entscheidend, daß das Laufverhalten von DNA Fragmenten in der Elektrophorese nicht nur von der Länge, sondern auch von der Zusammensetzung der Fragmente abhängt. Dies gilt insbesondere für hochauflösende Molekulargewichtsstandards in denen es nur sehr geringe Längenabstände gibt.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher die Verwendung
20 informationstragender Polymere, insbesondere der erfindungsgemäß erhältlichen informationstragenden Polymere, zur Herstellung von Molekulargewichtsstandards.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Herstellung
25 polymerer Datenspeicher

Die beschriebenen Logomere können als RAM (Lesen und Schreiben) und ROM (nur Lesen) Datenspeicher hoher Kapazität und niedrigen Energieverbrauchs genutzt werden. Um eine möglichst hohe Informationsdichte zu gewährleisten und die Handhabung des Speichers
30 möglichst einfach zu machen, werden die Logomere an einen festen Träger gebunden. Dies gewährleistet, daß die Logomere einzeln adressiert, d.h. einzeln geschrieben, gelesen und gelöscht werden können (siehe Abbildung 9).

Die benötigten Schreib- und Leseoperationen werden auf der Basis der im erfindungsgemäßen Schritt V und der unter *Auslesen der in Logomeren enthaltenen*
35 *Information* beschriebenen enzymatischen Reaktionen durchgeführt, die ihrerseits durch Temperaturänderungen gesteuert werden. Die Temperaturänderungen können entweder per Laser oder durch Erhitzen und Abkühlen des festen Trägers z.B. in einem Thermozykler durchgeführt werden.

Zum Schreiben wird eine Modifikation des oben beschriebenen Verfahrens der
40 Symbolpolymerisation verwendet. Die Polymerisation findet an einem festen Träger statt, die Algomere werden durch wiederholte Restriktions-Ligationszyklen miteinander verkettet.

Um einzelne, adressierbare Speicherzellen herzustellen, werden in bestimmten Abständen einzelsträngige "Ankermoleküle" irreversibel (durch UV-Bestrahlung oder im Ofen) an den Träger gebunden. Der solcherart mit Ankermolekülen versehene Träger kann nun reversibel mit Logomeren beschrieben werden, die in einem Löschvorgang wieder davon entfernt werden können.

Bei einem Schreibvorgang werden zuerst spezifische Algomere ("Starter-Algomere") durch Hybridisierung an die Ankermoleküle gebunden. Diese Starter-Algomere sind ihrerseits Ausgangspunkt einer Symbolpolymerisation, bei der weitere Algomere miteinander verkettet werden. Die Verkettung der Algomere erfolgt dabei, abweichend von dem im Schritt V des erfindungsgemäßen Verfahrens beschriebenen Vorgehen, durch wiederholte Zyklen von Restriktion und Ligation. Bei jedem Zyklus wird das jeweils letzte Algomer einer polymerisierenden Kette mit einem Restriktionsenzym geschnitten, so daß eine einzelsträngige Überhangsequenz (Elongationspunkt) entsteht, an die seinerseits das nächstfolgende Algomer durch Ligation gebunden werden kann.

Für den solcherart beschriebenen Schreibvorgang müssen die Algomere ein spezifisches Design aufweisen:

- a) Jedes zu schreibende Algomer besitzt eine einzelsträngige Überhangsequenz, mit der es an den Elongationspunkt eines Logomers binden kann.
- b) Ein Algomer, das nur eine einzelsträngige Überhangsequenz besitzt mit der es an einen Elongationspunkt binden kann, selbst aber keine Restriktionsschnittstelle besitzt, durch die es von Restriktionsenzymen erkannt werden kann und damit nicht als Elongationspunkt dienen kann, heißt Ende-Algomer.
- c) Jedes zu schreibende Algomer, das nicht das Ende-Algomer ist, besitzt ein doppelsträngiges Ende, worin sich die Erkennungssequenz für das jeweils gewählte Restriktionsenzym befindet. Bei einer Restriktion spaltet das Enzym das Algomer, so daß eine einzelsträngige Überhangsequenz entsteht, die ihrerseits selbst wieder als Elongationspunkt dienen kann.
- d) Jedes fertige Logomer besitzt genau ein Starter- und ein Ende-Algomer. Starter- und Ende-Algomere sind, entsprechend der oben gegebenen Definition, *Terminatoren*.
- e) Die Überhangsequenzen der Algomere sind kompatibel zu einem jeweils gewählten Restriktionsenzym, so daß ein zu verkettendes Algomer an den Elongationspunkt eines Logomers binden kann und der nachfolgend durch Restriktion entstehende Elongationspunkt identisch zu seinem Vorläufer ist.
- f) Die im Algomer auf die Überhangsequenz folgende Sequenz ist solcherart gewählt, daß ein verkettetes Algomer durch einen nachfolgenden Restriktionsvorgang nicht wieder abgespalten werden kann.

Als fester Träger eignen sich Membranen wie Gene Screen Plus (DuPont, Biotechnology Systems, Katalog Nr.: NEF-986 oder NEF-987), Hybond-N (Amersham Life Sciences, Katalog

Nr.: RPN 82N oder RPN 137N), Silizium-, Silicat-Oberflächen oder Glas. Die Ankermoleküle werden jeweils kovalent daran gebunden (UV-Bestrahlung oder Hitze).

Damit die Moleküle einzeln adressierbar sind, müssen sie in regelmäßigen Abständen auf dem Träger angebracht werden.

- 5 Als Restriktionsenzyme eignen sich handelsübliche Enzyme, vorzugsweise solche, die nicht bei 65°C deaktiviert werden (z.B. HindIII).

Die Schreiboperationen basieren darauf, daß die zum Schreiben verwendeten Enzyme unterschiedliche Temperaturoptima haben. So liegt das Temperaturoptimum der zum Verketteten von Symbolen verwendeten Ligase bei 16°C, wohingegen die benötigten

- 10 Restriktionsenzyme nur bei 37°C funktionieren. Dadurch ist das getaktete Schreiben von Symbolen in Zyklen von Restriktion und Ligation möglich.

Da alle Lese- und Schreiboperationen mithilfe von Enzymen durchgeführt werden, ist es nötig, den Träger temperieren zu können. Dabei können, bei der Verwendung eines Thermozyklers und thermomanipulierbaren Materials, alle Speicherzellen simultan derselben Operation
15 unterworfen werden. Alternativ können Speicherzellen einzeln und unabhängig voneinander zugegriffen werden, wenn ein entsprechend kleiner Schreib-/Lesekopf eingesetzt wird. Dieser Schreib-/Lesekopf muß einerseits als Pipettiervorrichtung für die benötigten Moleküle (Enzyme, Algomere) dienen und andererseits die für eine bestimmte Manipulation benötigte Temperatur erzeugen (z.B. mit Hilfe eines Lasers).

- 20 Beim Löschvorgang wird ein Logomer durch Denaturierung vom jeweiligen Ankermolekül gelöst. Das Ankermolekül steht danach wieder für einen Schreibvorgang zur Verfügung. Für einen Lesevorgang wird ein Logomer durch Denaturierung von einem Ankermolekül gelöst. Es kann dann nach den erfindungsgemäßen Verfahren ausgelesen werden.

- Die beschriebenen Operationen Schreiben, Löschen, Lesen werden durch Enzyme bei
25 verschiedenen Temperaturen durchgeführt. Dabei können, bei der Verwendung eines thermomanipulierbaren Materials, alle Speicherzellen simultan derselben Operation unterworfen werden. Alternativ können Speicherzellen einzeln und unabhängig voneinander zugegriffen werden, wenn ein entsprechend kleiner Schreib-/Lesekopf eingesetzt wird. Dieser Schreib-/Lesekopf muß einerseits als Pipettiervorrichtung dienen und andererseits die für eine
30 bestimmte Manipulation benötigte Temperatur erzeugen.

- Der Vorteil des hier vorgestellten Polymerspeichers besteht gegenüber den bisher verwendeten Materialien in einer höheren Speicherdichte (DNA Moleküle haben eine Größe, die bei 10^{-10} m liegt), und einem sehr viel geringeren Energieverbrauch (zur Berechnung der in
35 enzymatischen Reaktionen benötigten Energie siehe [L. M. Adleman, Molecular Computation of Solutions to Combinatorial Problems, *Science*, 266, 1021-1024, (1994)]). Außerdem ergibt sich eine erheblich verbesserte Umweltverträglichkeit, weil die hier beschriebenen Polymere Biomoleküle ohne jegliche Toxizität sind.

- Der hier beschriebene Ansatz hat gegenüber den bisher für das DNA Computing verwendeten
40 wässrigen Lösungen den Vorzug höherer Speicherdichte und der Adressierbarkeit einzelner Polymere.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher ein polymerer Datenspeicher, umfassend erfindungsgemäße informationstragende Polymere, erfindungsgemäße Verfahren zur Verknüpfung von Molekülen, erfindungsgemäße Ausleseverfahren oder
5 erfindungsgemäße Isolierungs- und Vervielfältigungsverfahren.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Herstellung eines DNA Computers

10

Mithilfe der bereits beschriebenen Informationsdarstellung durch Algomere und Logomere kann ein DNA Computer konstruiert werden. Der DNA Computer umfaßt (s. Abbildung 10):

- a) Oligonukleotidsynthesizer
 - b) Thermozykler
 - 15 c) Pipettiervorrichtung
 - d) Vorrichtung zur Isolation von Polymeren
 - e) Vorrichtung zur Auftrennung von Nukleinsäuren, z.B. Elektrophoresesystem
 - f) Scanner
 - g) Steuerungsrechner (Workstation, z.B. PC)
- 20 Der DNA Computer ist konstruiert als ein Hybridsystem, bei dem einfache, rechenintensive Algorithmen oder die Speicherung von Daten mithilfe von DNA und den dazu benötigten Geräten (Geräte a) bis e)) vorgenommen werden und ein herkömmlicher Computer (vorzugsweise Workstation, z.B. PC) als Host und Steuerungsrechner dient.
- Der DNA Computers beinhaltet mindestens einen Thermozykler, der als „Informationsreaktor“
25 dient und über den Steuerungsrechner gesteuert werden kann (z.B. MJ Research PTC-100, Eppendorf Mastercycler). In den Tubes oder Mikrotiterplatten des Thermozyklers befinden sich Molekülgemische aus Nukleinsäuren, Enzymen und weiteren chemischen Substanzen (z.B. Puffer, Nukleotide), die für die Algomer-Assemblierung (Schritt IV des erfindungsgemäßen Verfahrens) und die Symbolpolymerisation (Schritt V des
30 erfindungsgemäßen Verfahrens) benötigt werden. Programmiert wird der DNA Computer durch die Implementierung von Algorithmen als Grammatiken (Schritt I des erfindungsgemäßen Verfahrens). Die für die Grammatiken jeweils benötigten Monomersequenzen werden durch das NFR-Verfahren und die Benutzung eines Oligonukleotidsynthesizers (z.B. ABI 392, ABI 398, ABI 3948 von Perkin-Elmer Applied
35 Biosystems) hergestellt (Schritte II und III des erfindungsgemäßen Verfahrens). Diese Monomersequenzen gehen als Input in den Thermozykler ein, in dem die Algomer-Assemblierung (Schritt IV des erfindungsgemäßen Verfahrens) stattfindet.
- Die so gewonnenen Algomere werden in eine oder mehrere Reaktionskammern des Thermozyklers überführt, wo sie zu Logomeren verknüpft werden (Symbolpolymerisation,
40 Schritt V des erfindungsgemäßen Verfahrens). Die so gewonnen Logomere werden in einer eigenen Vorrichtung isoliert und vervielfältigt. Diese Vorrichtung kann nach den Prinzipien

arbeiten, die in dem erfindungsgemäßen Verfahren beschrieben worden sind. Nach Isolierung und Vervielfältigung können die Logomere gelesen werden. Dazu werden sie wiederum in einem Thermozykler erfindungsgemäß ausgelesen. Die aus dem Ausleseverfahren resultierenden Nukleinsäurefragmente können, wie bereits beschrieben, durch
5 Gelelektrophorese gelesen werden. Hier dient das Gelelektrophoresegerät als Ausgabegerät des DNA Computers, die solcherart erhaltenen Bandenmuster werden von einem Scanner in den Steuerungsrechner eingelesen.

Der Steuerungsrechner nimmt jetzt seinerseits aufgrund der solcherart erhaltenen Ergebnisse weitere Berechnungen und Steuerungen vor und setzt ggfs. weitere molekulare Reaktionen in
10 Gang. Z.B. kann der Steuerungscomputer aufgrund erhaltener Ergebnisse die Neuimplementierung von Grammatiken (Herstellung von Oligomeren mit dem NFR-Verfahren) veranlassen. Auf diese Weise können die durch molekulare Verfahren als Logomere erhaltenen Informationen nicht nur als passive Daten sondern auch als Instruktionen (und Adressen) dienen. Dadurch ist ein sich selbst steuernder Prozeß
15 verschiedener Algorithmen möglich, der für eine universelle Datenverarbeitung nötig ist.

Der DNA Computer verarbeitet dabei Informationen als Oligo- und Polymere. Als Speicherzellen der Daten dienen Klonierungsvektoren (Plasmide), die sich in flüssiger Lösung oder auf einem festen, adressierbaren Träger befinden.

Der DNA Computer ist wie oben beschrieben mittels Grammatiken programmierbar und kann
20 von einem herkömmlichen Computer, der als Steuerungsrechner und Host-System fungiert, gesteuert werden.

Da der DNA Computer bestimmte Operationen in molaren Größenordnungen (z.B. die Symbolpolymerisation, wie in Schritt V des erfindungsgemäßen Verfahrens beschrieben, in 10^{12} - 10^{20} Elementaroperationen pro Sekunde) vornehmen kann, ist eine mögliche
25 Anwendung die Berechnung einfacher, rechenintensiver Algorithmen. Jedoch kann der DNA Computer auch für andere Anwendungen, z.B. die Erzeugung bestimmter, regelmäßiger DNA Strukturen, die z.B. innerhalb der Nanotechnologie interessant sind, verwendet werden.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher ein DNA-Computer, umfassend
30 erfindungsgemäße informationstragende Polymere. Ein weiterer Gegenstand der vorliegenden Erfindung ist ein DNA-Computer, in dem ein erfindungsgemäßes Ausleseverfahren und/oder erfindungsgemäße Isolierungs- und Vervielfältigungsverfahren eingesetzt werden.

35

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Herstellung kleinster molekularer Strukturen mit Logomeren

- 5 Mit den hier beschriebenen Logomeren lassen sich kleinste molekulare Strukturen (Nanostrukturen) herstellen. Bausteine solcher Strukturen sind Algomere, die zu regelmäßigen Mustern von Logomeren zusammengefügt werden können. Z.B. können verschiedene Algomere mit unterschiedlichen Fremdmolekülen (Liganden) dotiert werden, um regelmäßige, höhermolekulare Muster der jeweiligen Liganden zu erhalten. Die Logomere fungieren dabei als "Skelett" molekularer Baugruppen von Liganden.
- 10 Z.B. könnte ein binäres Logomer alternierende Muster leitfähiger und nicht-leitfähiger Liganden enthalten. Durch die Anordnung vieler Logomere auf einem festen Träger können so Leiterbahnen im Nanometerbereich entstehen.
- Als Liganden können aber auch Biomoleküle, Antikörper, optisch aktive Moleküle o.a. verwendet werden.

- 15 Logomere können als "intelligenter" Klebstoff eingesetzt werden (siehe Abbildung 18). Dabei macht man sich die Hybridisierung komplementärer Nukleotide zunutze. Bringt man auf die zu verklebenden Flächen Logomere auf, die etwa kovalent an die zu verklebenden Flächen gebunden werden, so können ansonsten völlig identische Flächen nur dann verkleben, wenn
- 20 sie komplementäre Sequenzen aufweisen, da nur jeweils Flächen komplementärer Nukleinsäuren Wasserstoffbrückenbindungen ausbilden. Die "Intelligenz" des Klebstoffes besteht also in einer durch die jeweils verwendeten Sequenzen hochselektiven Adhäsionswirkung. Dabei kann auch die Adhäsionsstärke für zu verklebende Flächen genau eingestellt werden: Zum einen kann über die Dichte der Logomere auf den jeweiligen Flächen
- 25 deren Adhäsionsstärke eingestellt werden, zum anderen kann über das AT/GC Verhältnis der Logomere deren Schmelzpunkt variiert werden, so daß verklebte Flächen bei unterschiedlichen Temperaturen ihre Adhäsion verlieren. Aufgrund der Tatsache, daß hier unschädliche, leicht abbaubare Biomoleküle als Klebstoff verwendet werden, können derartige Klebstoffe auch zur medizinische Verwendung (z.B. Chirurgie, Mikrochirurgie) eingesetzt
- 30 werden. Eine andere Verwendung ist die Nutzung des beschriebenen Verfahrens in Hochpräzisions- und Authentifizierungsanwendungen.

- Versetzt man die zur Herstellung von Logomeren verwendeten Algomere mit Sequenzen zur Bindung von Liganden, so können stärkere Adhäsionswirkungen erreicht werden. Solche Liganden können im Falle von DNA DNA-bindenden Proteinen, z.B. spezifische, gegen DNA-
- 35 Sequenzen gerichtete Antikörper sein. Diese können z.B. jeweils über einen weiteren Liganden untereinander verbunden werden (siehe Abbildung 19). Eine einfache, nicht-programmierbare Adhäsionswirkung kann auch ohne Logomere erreicht werden, wenn die verwendeten Proteine (z.B. Antikörper) ihrerseits an die zu verklebenden Flächen binden können (wie z.B. Antikörper, wenn die zu verklebende Fläche ihr Antigen ist). Siehe dazu auch
- 40 Abbildung 20. Einfachere Klebstoffe sind möglich, wenn, an bestimmte Flächen spezifisch bindende, Proteine gentechnisch hergestellt werden.

Eine alternative Methode besteht darin, die zu verklebenden Flächen nicht über schwache Wechselwirkung, sondern über kovalente Bindung miteinander zu verbinden. Abweichend vom oben vorgestellten Verfahren wird die Adhäsionskraft über die Dichte von Algomeren eingestellt. Dabei werden die zu verklebenden Flächen mit Algomeren, die zueinander jeweils komplementäre sticky ends tragen versehen. Diese können dann hybridisieren und durch Ligation miteinander kovalent verbunden werden.

Mit Logomeren können Oberflächen unterschiedlichster Struktur und unterschiedlichster physikalischer Eigenschaften hergestellt werden. Den Anwendungen ist gemein, daß sich mit den Logomeren nahezu beliebige Muster im Nanometerbereich bilden lassen, die das Skelett kleinster molekularer Bauelemente sein können.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher die Verwendung informationstragender Polymere; insbesondere der erfindungsgemäß erhältlichen informationstragenden Polymere, zur Herstellung oder Bearbeitung kleinster molekularer Strukturen oder als molekularer Kleber.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Herstellung nanotechnologischer Baukastensysteme.

Mit dem oben beschriebenen NFR-Verfahren und dem "Parallel Extension" Verfahren können erstmals größere Grammatiken direkt in Moleküle übersetzt werden. Dabei sind diese Grammatiken nicht auf reguläre Grammatiken beschränkt. U. a. aufgrund der Wiederverwendbarkeit von Sequenzen lassen sich damit z.B. auch ψ -Moleküle [Matthias Scheffler, Axel Dorenbeck, Stefan Jordan, Michael Wüstefeld, Günter von Kiedrowski, Self-Assembly of Trisigonucleotidyls: The Case for Nano-Acetylene and Nano-Cyclobutadiene. *Angewandte Chemie Int. Ed.*, 38(22), 3312-3315, (1999)] und DX-Moleküle [Erik Winfree, Furong Liu, Lisa A. Wenzler & Nadrian C. Seeman, Design and self-assembly of two-dimensional DNA crystals, *Nature*, 394, 539-544, (1998)] programmieren.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher die Verwendung des NFR-Verfahrens, des "Parallel-Extension" Verfahrens und der Übersetzung von Grammatiken in Moleküle zur Herstellung von Komponenten nanotechnologischer Baukastensysteme.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die kontrollierte, programmierbare Herstellung von biologisch aktiven Nukleinsäuren.

Dazu setzt man z.B. als Terminale anstelle künstlicher Sequenzen zur Darstellung von Symbolen biologisch aktive Sequenzen wie Restriktionssites, Rekombinationssites, Centromere, Promotoren, Exons, Gene o.a. ein. Diese können dann auf kontrollierte,

programmierbare Weise zu biologisch aktiven Konstrukten, z.B. zu Genen oder künstlichen Chromosomen zusammengesetzt werden.

- 5 Ein weiterer Gegenstand der vorliegenden Erfindung ist daher die Verwendung informationstragender Polymere, insbesondere der erfindungsgemäß erhältlichen informationstragenden Polymere, zur kontrollierten Herstellung biologisch aktiver Nukleinsäuren.

Referenzen

Patente:

International Publication Number	Inventor	Title
US4683202	Kary B. Mullis, Kensington, Calif.	Process for amplifying nucleic acid sequences
WO 97/07440	Adleman, Leonard, M; 18262 Hastings Way, Northridge, CA 91326 (US)	MOLECULAR COMPUTER
WO 97/29117	GUARNIERI, Frank [US/US]; 62 Lake Street, Brooklyn, NY 11223 (US). BANCROFT, Frank, Carter [US/US]; 51 Dewey Street, Huntington Station, NY 11743 (US).	A DNA-BASED COMPUTER
US5804373	Schweitzer; Allan Lee, Plainsboro, NJ Smith; Warren D. , Plainsboro, NJ	Molecular automata utilizing single- or double-strand oligonucleotides

Veröffentlichungen:

L. M. Adleman, Molecular Computation of Solutions to Combinatorial Problems, *Science*, **266**, 1021-1024, (1994)

Breslauer, K.J., Frank, R., Blocker, H., Marky, L.A., *Proc. Natl. Acad. Sci.*, **83**, 3746-3750, 1989

Capecchi M.R., Altering the genome by homologous recombination. *Science*, **244**(4910), 1288-1292, (1989)

Chomsky, N., Three models for the description of language, *JACM*, **2:3**, 113-124, (1956)

Chomsky, N., On certain formal properties of grammars, *Inf. and Control*, **2:2**, 137-167, (1959)

Chomsky, N., Formal properties of grammars, *Handbook of Math. Psych.*, **2**, 323-418, (1963)

Frank Guarnieri, Makiko Fliss, Carter Bancroft, Making DNA Add, *Science*, **273**, 220-223, (1996)

John E. Hopcroft, Formal Languages And Their Relation To Automata, (1969)

Jeffreys, Minisatellite repeat coding as a digital approach to DNA typing, *Nature*, **354**, 204-209, (1991)

Kilby, N. J., Snaith, M. R. & Murray, J. A., Site-specific recombinases: tools for genome engineering, *Trends Genet.*, **9**, 413-421, (1993)

Rolf Knippers, Molekulare Genetik, *Georg Thieme Verlag*, (1997)

Benjamin Lewin, Genes V, *Oxford University Press*, (1994)

Lund, A.H., Duch, M., Pedersen, F.S, Increased cloning efficiency by temperature-cycle ligation, *Nucleic Acids Research*, **24**: (4), 800-801, (1996)

J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)

Jens Niehaus, DNA Computing: Bewertung und Simulation, *Diplomarbeit am Fachbereich Informatik der Universität Dortmund, Lehrstuhl XI*, 116-123, (1998)

Qi Ouyang, Peter D. Kaplan, Shumao Liu, Albert Libchaber, DNA Solution of the Maximal Clique Problem, *Science*, **278**, 446-449, (1997)

Erik Winfree, Xiaoping Yang, Nadrian C. Seeman, Universal Computation via Self-assembly of DNA: Some Theory and Experiments, *Proceedings of the 2nd DIMACS Meeting on DNA Based Computers, Princeton University, June 20-12*, (1996)

Erik Winfree, Furong Liu, Lisa A. Wenzler & Nadrian C. Seeman, Design and self-assembly of two-dimensional DNA crystals, *Nature*, **394**, 539-544, (1998)

Matthias Scheffler, Axel Dorenbeck, Stefan Jordan, Michael Wüstefeld, Günter von Kiedrowski, Self-Assembly of Trisoligonucleotidyls: The Case for Nano-Acetylene and Nano-Cyclobutadiene. *Angewandte Chemie Int. Ed.*, **38**(22), 3312-3315, (1999)

Patentansprüche

1. Verfahren zur Herstellung von informationstragenden Polymeren, das umfaßt,
 - 5 I. eine reguläre Grammatik $G = (\Sigma, V, R, S)$ mit einem endlichen Terminalalphabet Σ , einer endlichen Variablenmenge V , einer endlichen Regelmenge R und einem Startsymbol S zu definieren;
 - II. das NFR-Verfahren (Niehaus-Feldkamp-Rauhe-Verfahren) zur Herstellung von Monomersequenzen;
 - 10 III. mit dem NFR-Verfahren eine in Schritt I definierte Grammatik zu implementieren, indem damit Monomersequenzen hergestellt werden, die die Regelmenge R einer Grammatik G eindeutig darstellen;
 - IV. aus den in Schritt III hergestellten Monomersequenzen für jede Regel der Regelmenge R von G ein die Regel repräsentierendes Oligomer zusammenzusetzen;
 - 15 V. die in Schritt IV zusammengesetzten Oligomere zu informationstragenden Polymeren zu verknüpfen.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das Terminalalphabet Σ einer Grammatik G die Terminale $0, 1, n$ Startterminale s (s_0, s_1, \dots, s_n) und m Endterminale e (e_0, e_1, \dots, e_m), enthält, wobei n und m jeweils größer oder gleich 0 sind und eine natürliche Zahl darstellen.
20
3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die in Schritt III konstruierte Monomersequenz Nukleotide, insbesondere Ribonukleotide, besonders bevorzugt Desoxyribonukleotide umfaßt.
25
4. Verfahren nach Anspruch 3, dadurch gekennzeichnet, daß die in Schritt III konstruierte Monomersequenz Proteinerkennungssequenzen (z.B. Restriktionsschnittstellen, Proteinbindungsstellen, Stopcodons) umfaßt.
30
5. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß man die Synthese der Monomersequenzen in Schritt II und III in vitro, vorzugsweise mittels eines Oligonukleotid-Synthesizers vornimmt.

6. Verfahren zur Isolierung und Vervielfältigung von informationstragenden Polymeren, die nach einem der vorhergehenden Ansprüche erhalten wurden, dadurch gekennzeichnet, daß man in Schritt V erhaltene informationstragende Polymere in Klonierungsvektoren ligiert, kompetente Zellen mit diesen Vektoren transformiert und die erfolgreich transformierten
5 Bakterien anhand von Selektionsmarkern selektioniert.

7. Verfahren zum Auslesen von Information aus informationstragenden Polymeren, die nach einem der vorhergehenden Ansprüche erhalten oder isoliert und vervielfältigt wurden, dadurch gekennzeichnet, daß man

- 10 a) mindestens n-1 das informationstragende Polymer enthaltende Lösungen mit jeweils einem Paar gegenläufiger Primer versetzt, wobei n für die Zahl der als Elongatoren in dem Polymer enthaltenen Oligomere steht;
- b) mindestens n-1 PCR-Ansätze durchführt, wobei n für die Zahl der als Elongatoren in dem Polymer enthaltenen Oligomere steht und jeweils ein Primer eines jeden Paares in dem dem Elongator gegenüberliegenden Terminator primt und der andere Primer
15 in dem Elongator selbst primt;
- c) die aus der PCR erhaltenen Polymer-Fragmente nach ihrer Länge durch Elektrophorese auftrennt und
- d) das aus der Elektrophorese erhaltene Muster optisch ausliest.

20 8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß das Auslesen in Schritt d) automatisiert durch einen Scanner oder eine Sequenziermaschine erfolgt.

9. Informationstragendes Polymer, erhalten nach einem der Ansprüche 1 bis 6.

25 10. Zufallszahlengenerator, umfassend informationstragende Polymere, insbesondere Polymere nach Anspruch 9.

11. Polymerer Datenspeicher, umfassend informationstragende Polymere nach Anspruch 9.

30 12. DNA-Computer, umfassend informationstragende Polymere nach Anspruch 9.

13. Biochip, umfassend informationstragende Polymere nach Anspruch 9.

14. Verwendung von informationstragenden Polymeren nach Anspruch 9 zur Herstellung von Molekulargewichtsstandards.
- 5 15. Verwendung von informationstragenden Polymeren nach Anspruch 9 zur Darstellung von Datenstrukturen.
16. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 als Marker oder Signaturen.
- 10 17. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Qualitätssicherung.
18. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Fälschungssicherung.
- 15 19. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung gentechnischer Erzeugnisse.
- 20 20. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung von Lebensmitteln.
21. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung von Organismen.
- 25 22. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung chemischer Erzeugnisse.
- 30 23. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung medizinischer und pharmazeutischer Erzeugnisse.

24. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung von Dokumenten.
- 5 25. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung von Geld.
26. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung von Gegenständen und Maschinen.
- 10 27. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung von Flüssigkeiten, Lösungen, Suspensionen oder Emulsionen.
- 15 28. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zur Verschlüsselung von Information.
29. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Authentisierung von Personen und Gegenständen.
- 20 30. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 als molekulare Kleber.
31. Verwendung von informationstragenden Polymeren nach Anspruch 9 zur Herstellung oder Bearbeitung kleinster molekularer Strukturen.
- 25 32. Verwendung von informationstragenden Polymeren nach Anspruch 9 zur Qualitätskontrolle synthetisch hergestellter Oligonukleotide.
- 30 33. Verwendung von informationstragenden Polymeren nach Anspruch 9 zur Herstellung von Biochips.

34. 1 bis n-bytige Nukleinsäuren, insbesondere 1 bis n-bytige Nukleinsäuren erhalten nach einem der Ansprüche 1 bis 6.
- 5 35. 1 bis n-bytige Biochips, insbesondere 1 bis n-bytige Biochips erhalten nach einem der Ansprüche 1 bis 6.
36. Verwendung von Biochips, insbesondere von Biochips nach einem der Ansprüche 13, 33 und 35 als Datenspeicher.
- 10 37. Verwendung von Biochips, insbesondere von Biochips nach einem der Ansprüche 13, 33 und 35 zur Herstellung von optischen Anzeigegeräten oder Bildschirmen.
38. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zur Kennzeichnung einzelner Moleküle, insbesondere von Nukleinsäuren, besonders bevorzugt von Genen.
- 15 39. Verwendung von Nukleinsäuren zur Verschlüsselung von Information.
40. Verwendung von Nukleinsäuren zur Verschlüsselung von Information, dadurch gekennzeichnet, daß zur Entschlüsselung kurze Nukleinsäuresequenzen (Primer) als Schlüssel benutzt werden.
- 20 41. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zur Verschlüsselung von Information, dadurch gekennzeichnet, daß das informationstragende Polymer in einer Vielzahl anderer Polymere verborgen ist.
- 25 42. Verwendung von Polymeren zum Zwecke der Kennzeichnung, dadurch gekennzeichnet, daß die Polymere verschlüsselt sind.
- 30 43. Herstellung biologisch aktiver Moleküle mit kontrolliertem, programmierbaren Verknüpfen von Sequenzen.

44. Herstellung biologisch aktiver Moleküle unter Verwendung von Algomenen.
45. Verfahren zum Auslesen von Information aus informationstragenden Polymeren die aus den Ansprüchen 1 bis 6 erhalten oder isoliert und vervielfältigt wurden, dadurch gekennzeichnet, daß man zum Auslesen Biochips benutzt.
46. Verwendung des NFR-Verfahrens zur Herstellung von Biochips.
47. Verwendung des NFR-Verfahrens zur Herstellung nanotechnologischer Komponenten oder von Komponenten nanotechnologischer Baukastensysteme.

Abbildungen

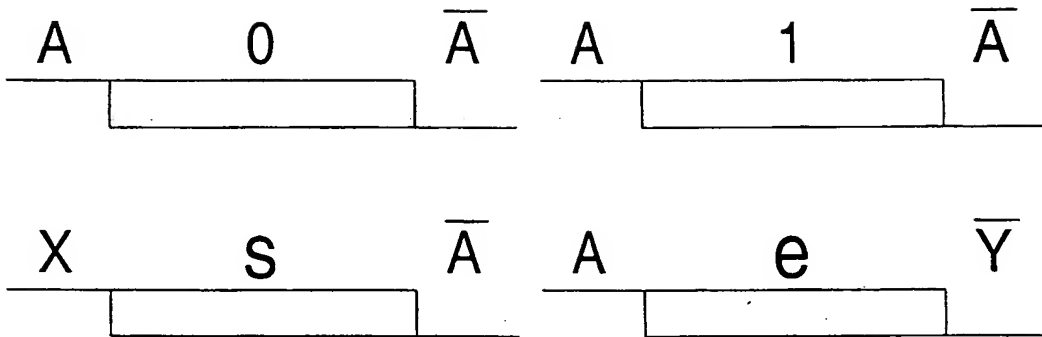


Abbildung 1

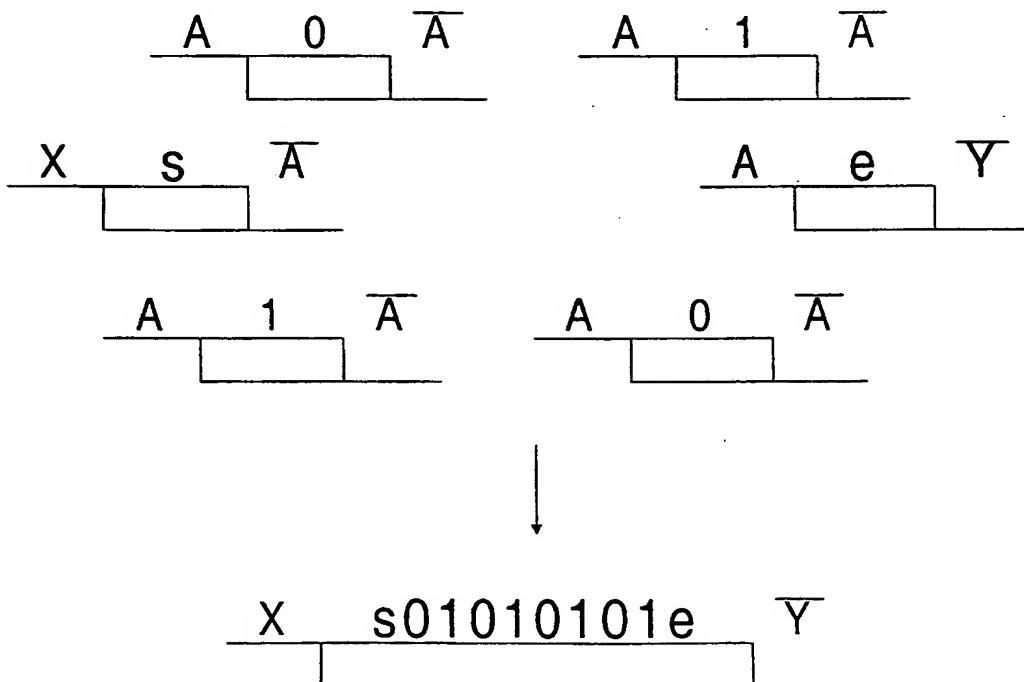


Abbildung 2

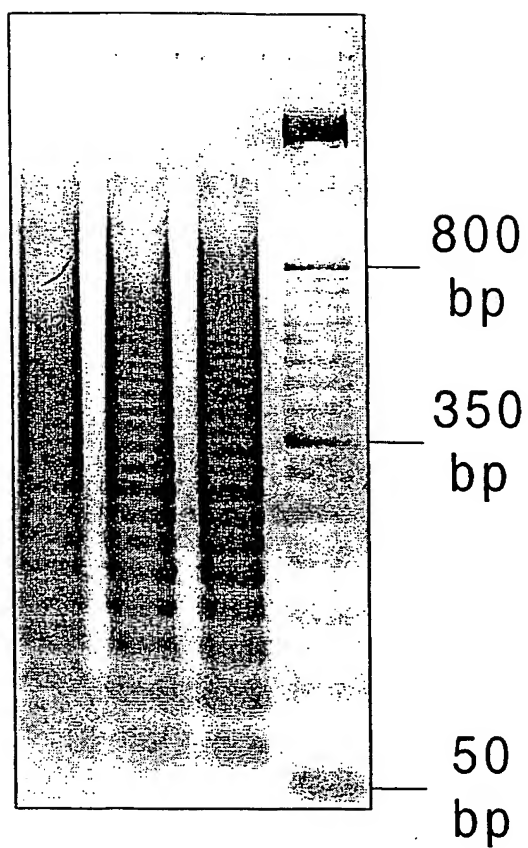


Abbildung 3

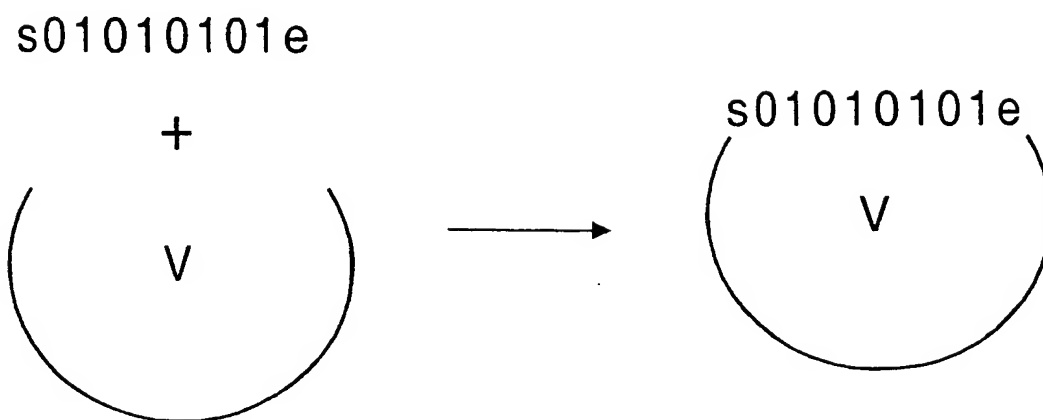


Abbildung 4

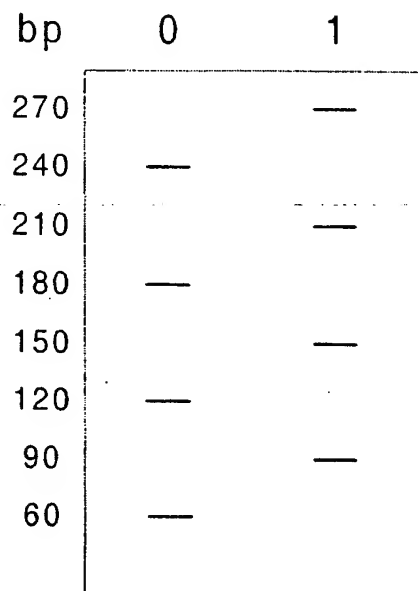
$\bar{s}\bar{o}\bar{1}\bar{0}\bar{1}\bar{0}\bar{1}\bar{0}\bar{1}e$ $\bar{s}\bar{o}\bar{1}\bar{0}\bar{1}\bar{0}\bar{1}\bar{0}\bar{1}e$ 

Abbildung 5

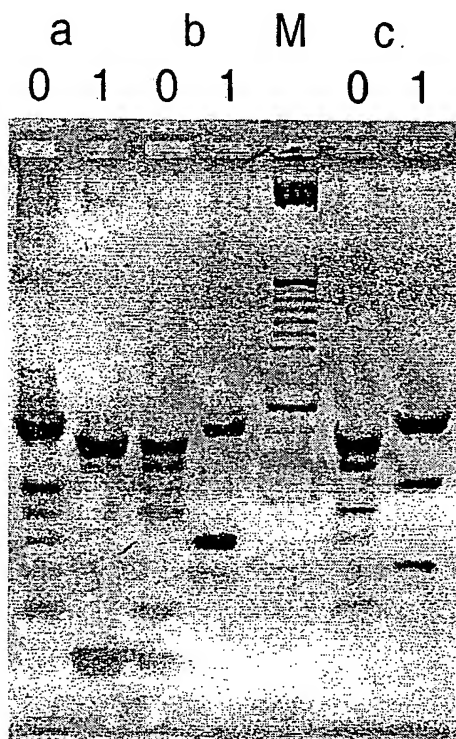


Abbildung 6

$\bar{s}\bar{o}\bar{1}\bar{0}\bar{1}\bar{0}\bar{1}\bar{0}\bar{1}e$ $\bar{s}\bar{o}\bar{1}\bar{0}\bar{1}\bar{0}\bar{1}\bar{0}\bar{1}e$

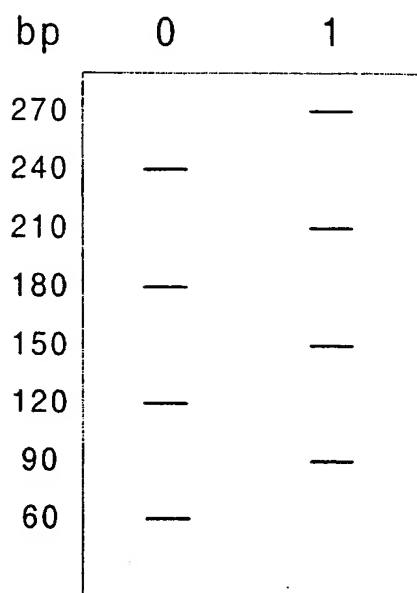


Abbildung 5

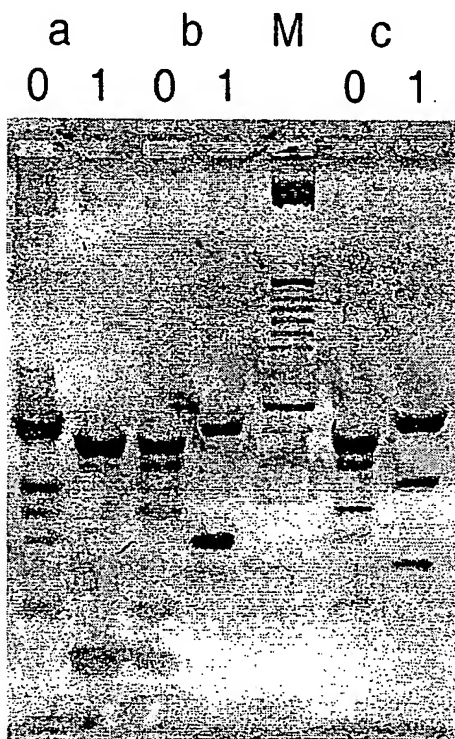


Abbildung 6

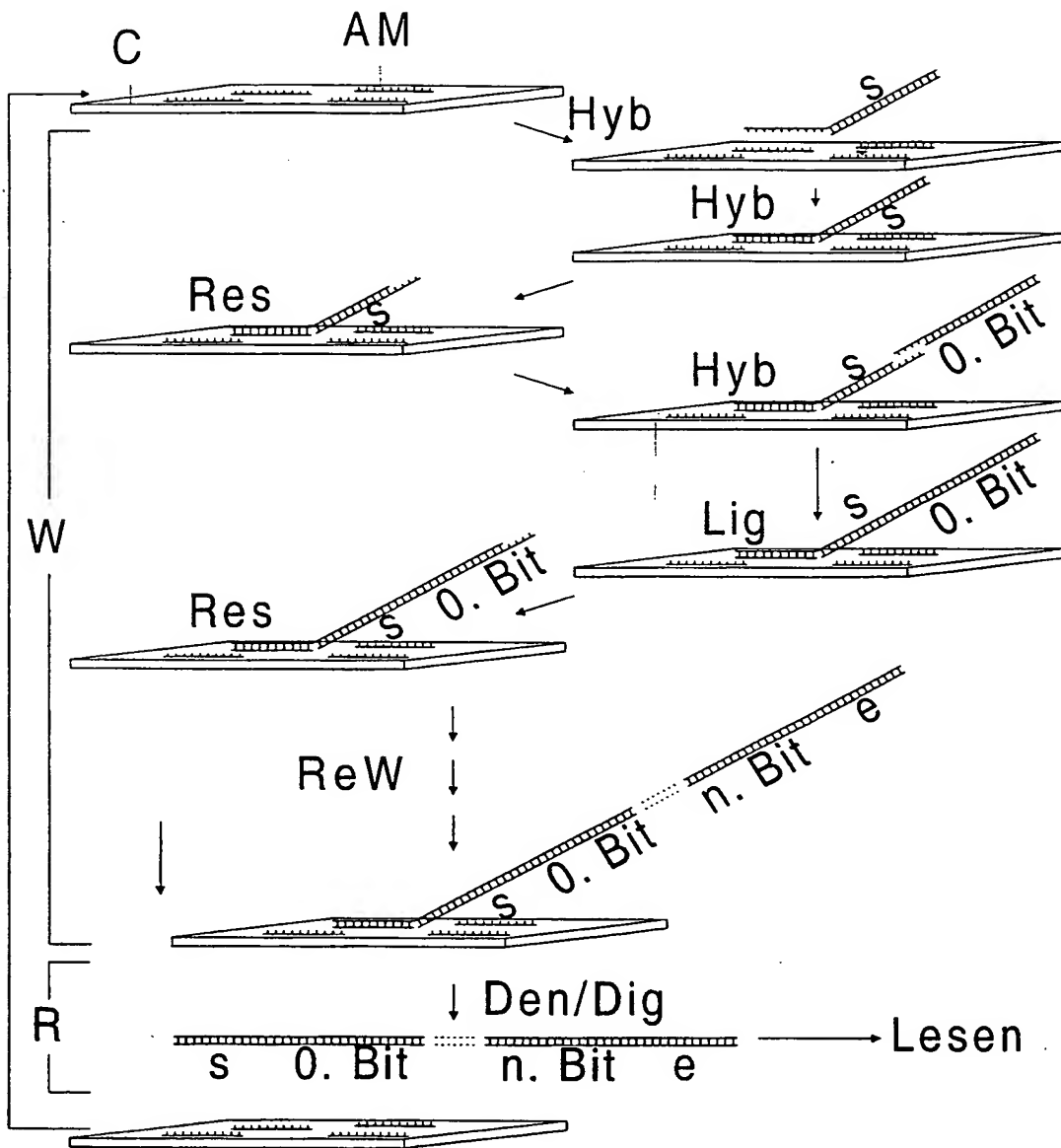


Abbildung 9

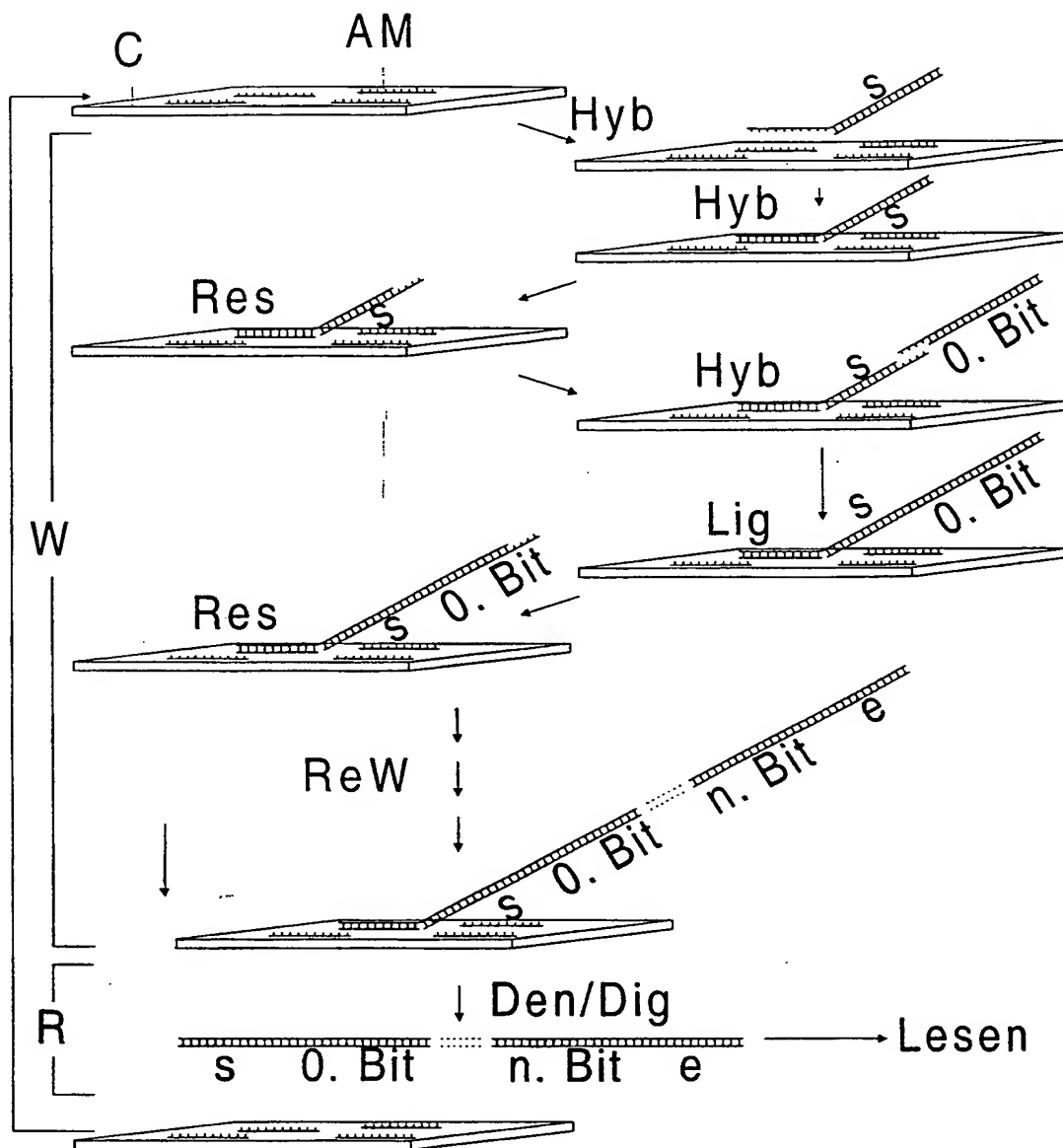
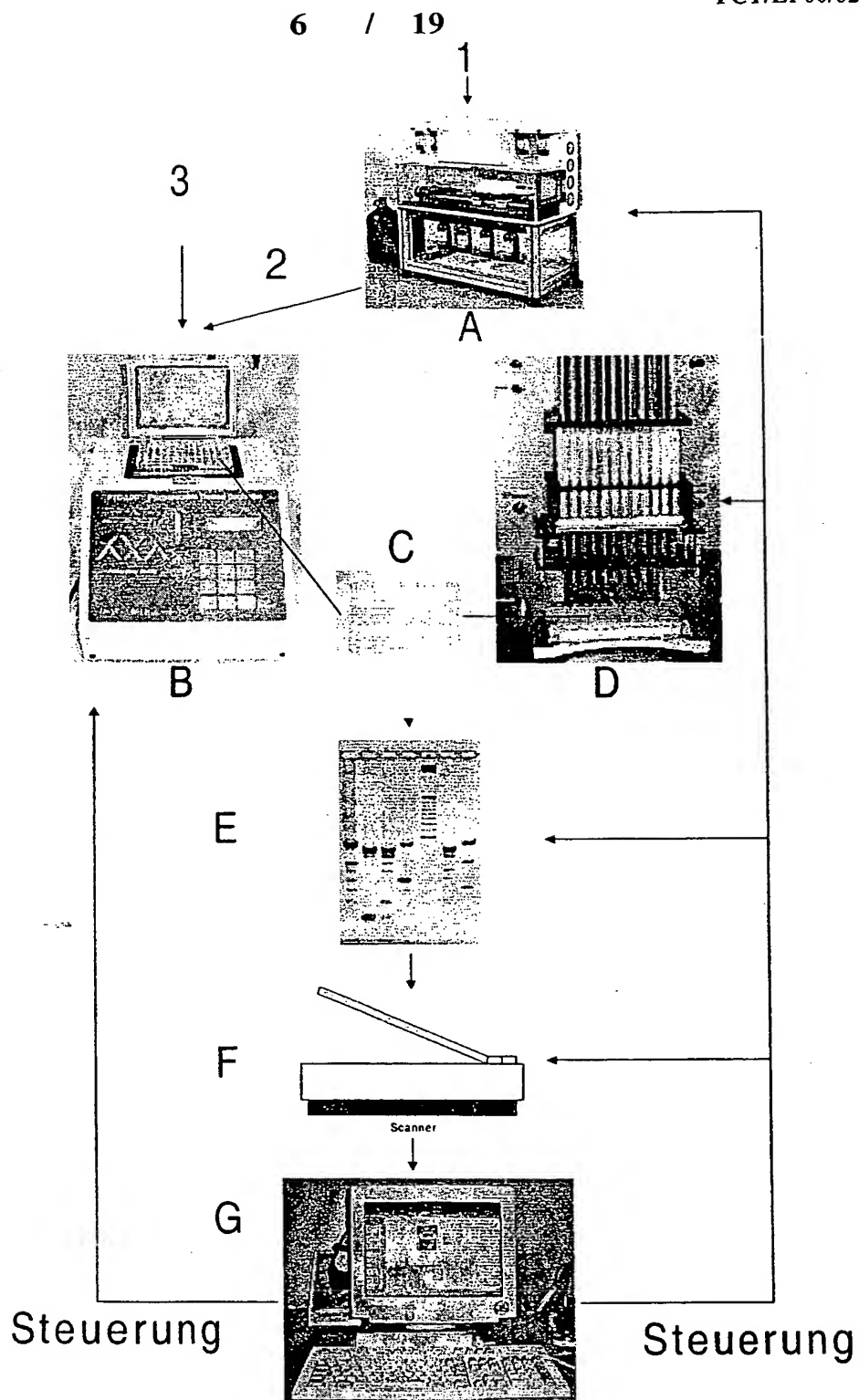


Abbildung 9



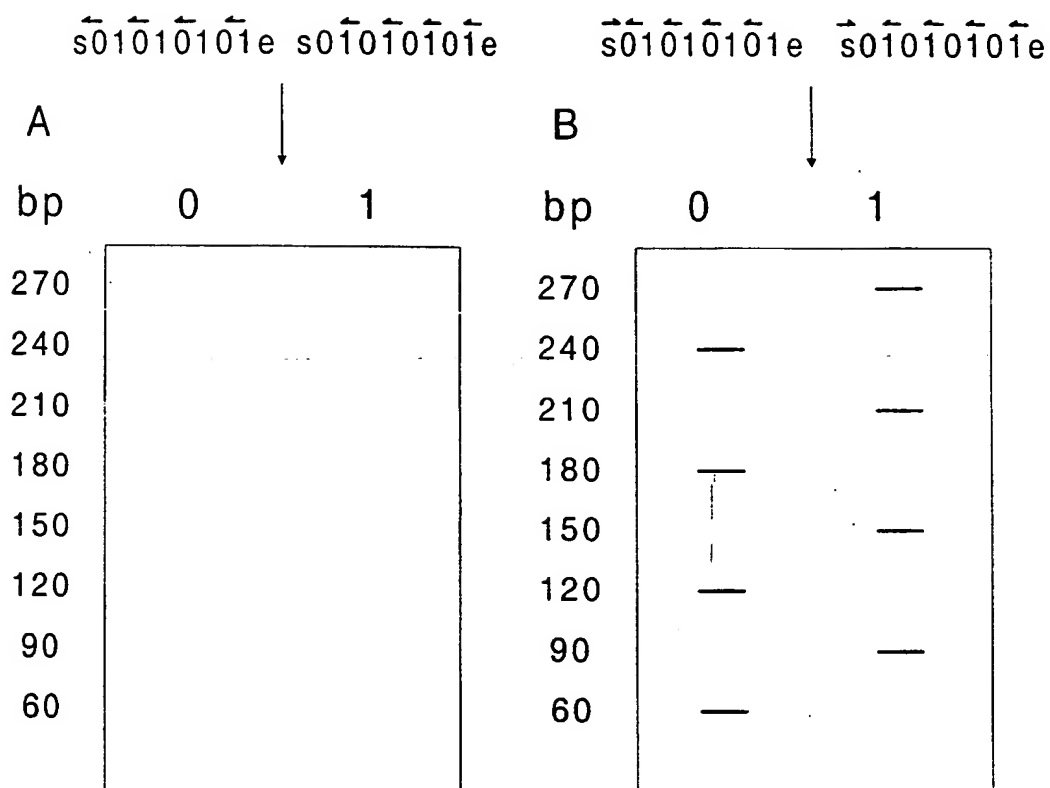


Abbildung 11

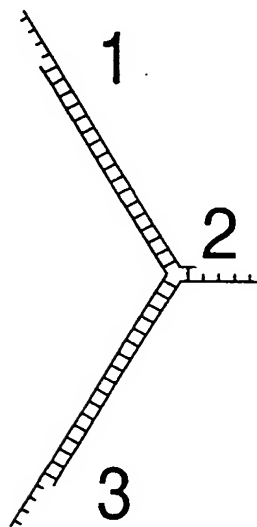


Abbildung 12

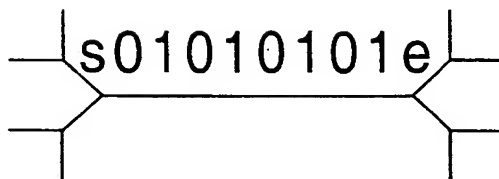


Abbildung 13

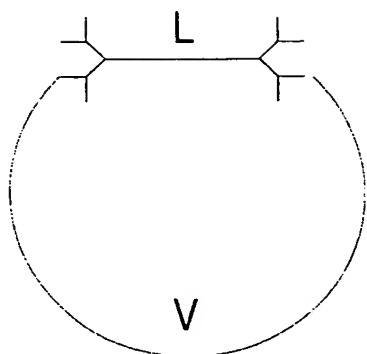


Abbildung 14

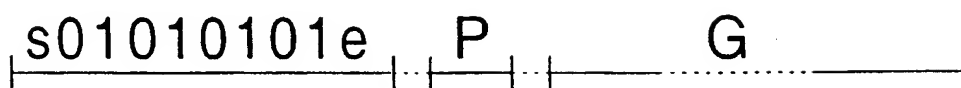


Abbildung 15

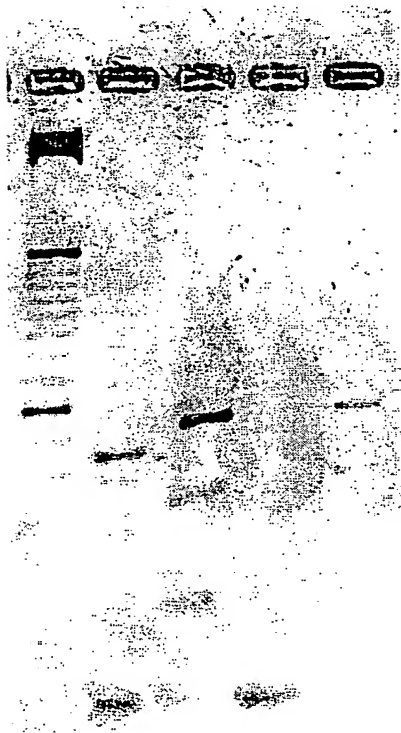


Abbildung 16

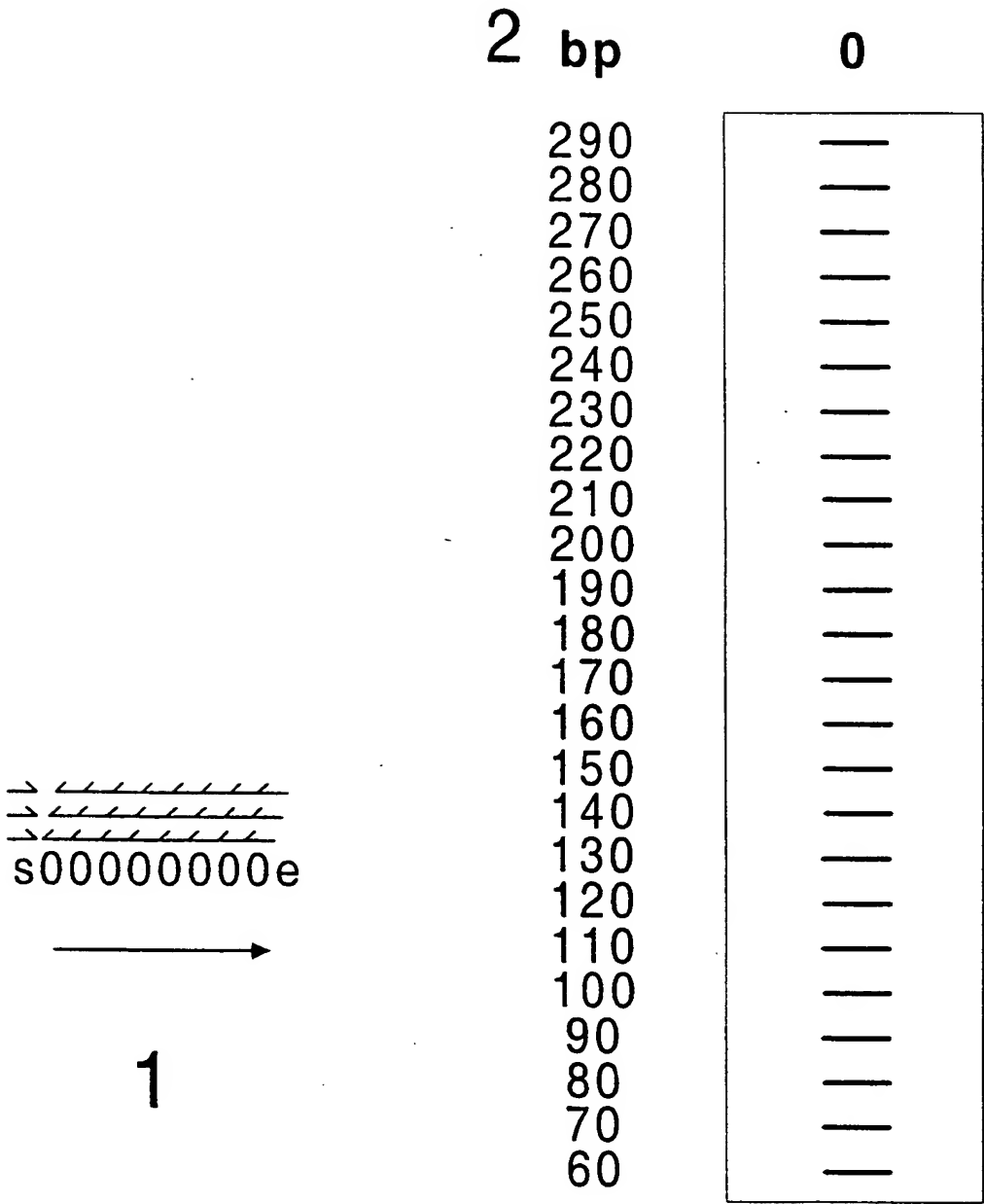


Abbildung 17

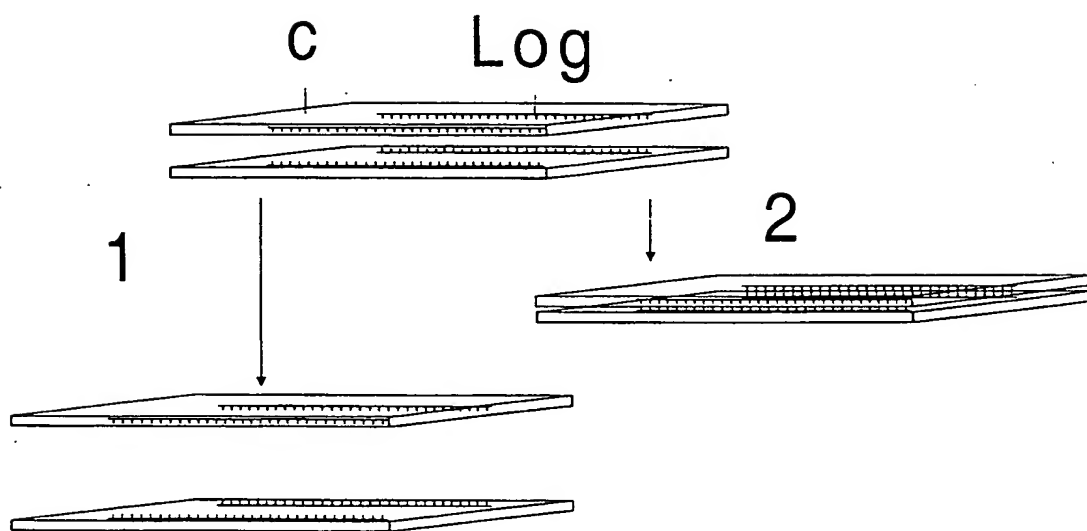


Abbildung 18

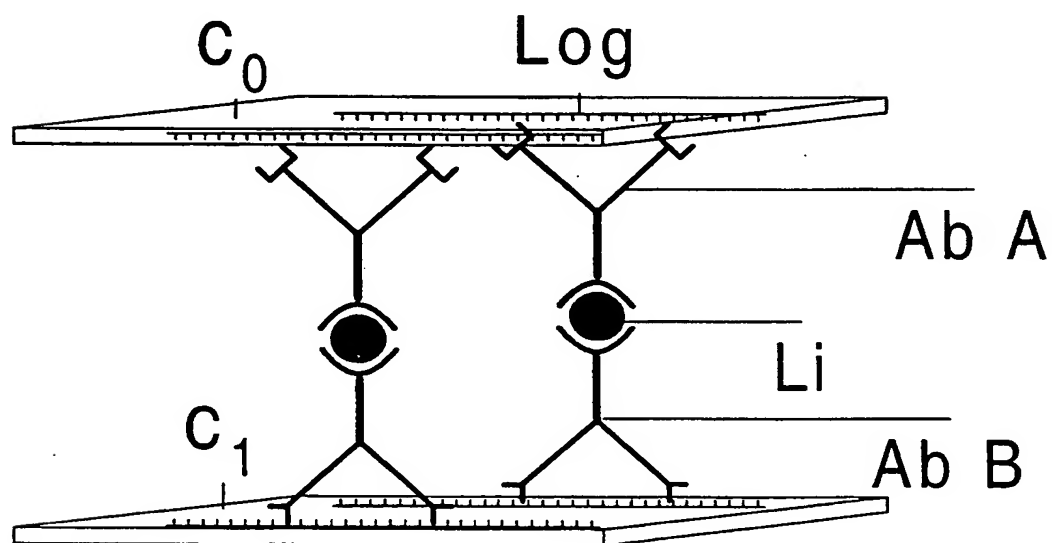


Abbildung 19

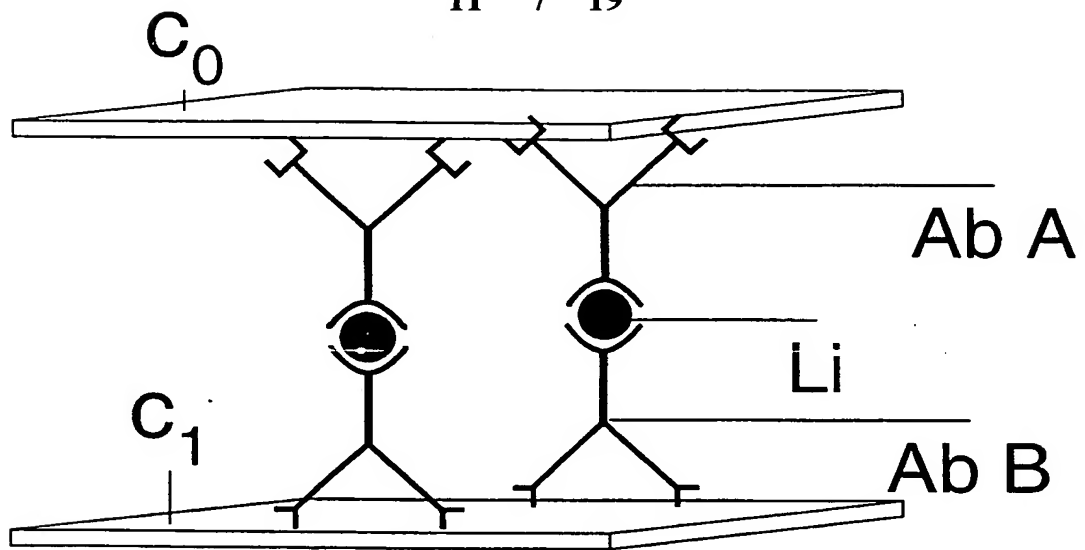


Abbildung 20

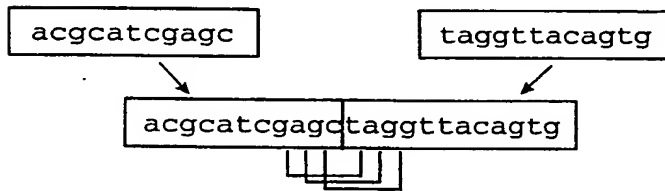


Abbildung 21

Regeln mit A:

S → aA

A → bA

A → cA

A → dB

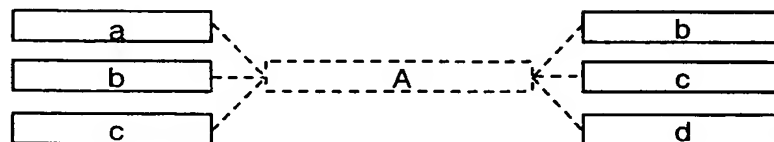


Abbildung 22

$R = \{ \dots, S \rightarrow 0A, S \rightarrow 0B, S \rightarrow 0C, S \rightarrow 0D, S \rightarrow 0E, \\ A \rightarrow 1F, B \rightarrow 2G, C \rightarrow 3H, D \rightarrow 4I, E \rightarrow 5J, \dots \}$

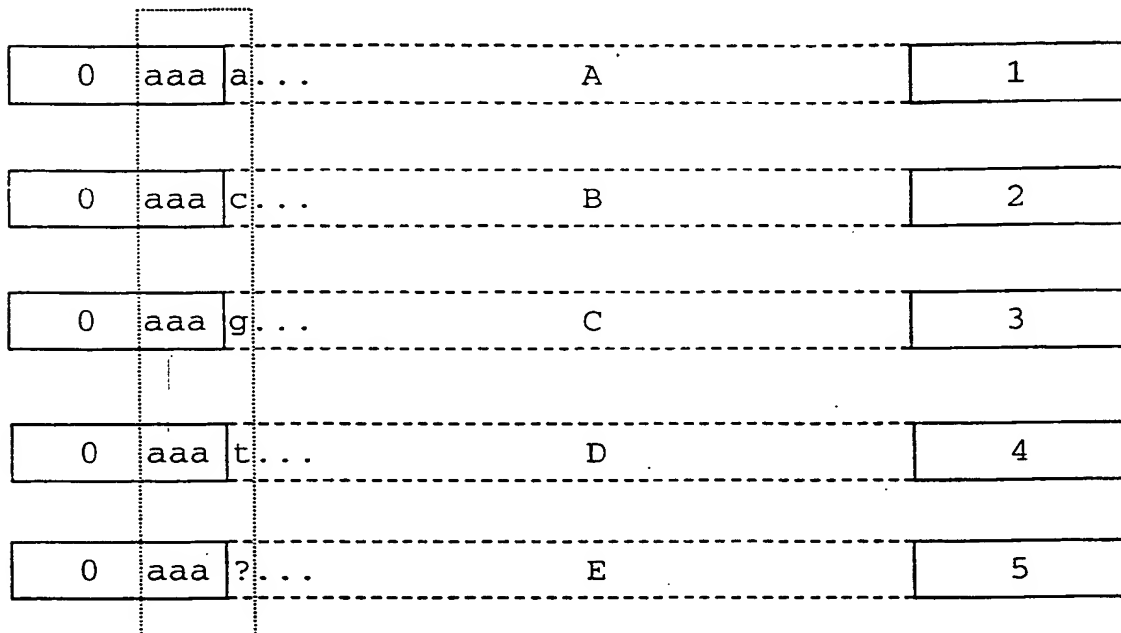


Abbildung 23

Rules:

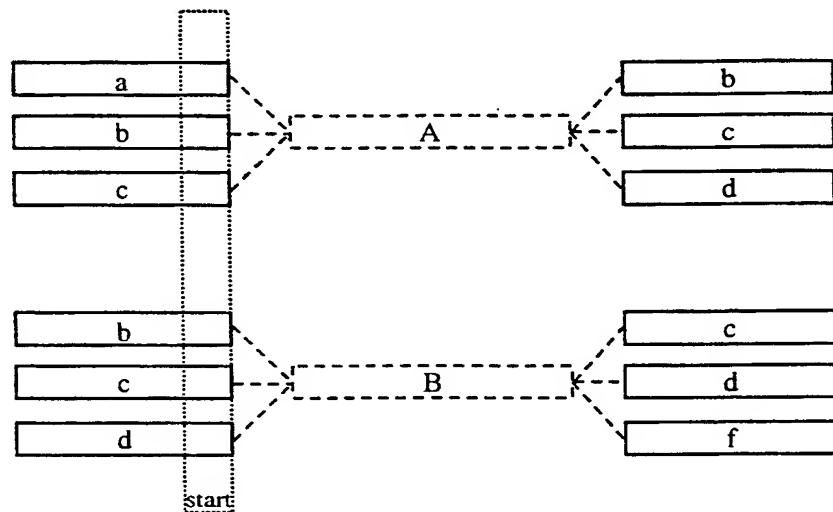
 $S \rightarrow aA$ $A \rightarrow bA$ $A \rightarrow cA$ $A \rightarrow dX$ $S \rightarrow bB$ $B \rightarrow cB$ $B \rightarrow dB$ $B \rightarrow fX$ 

Abbildung 24



Abbildung 25

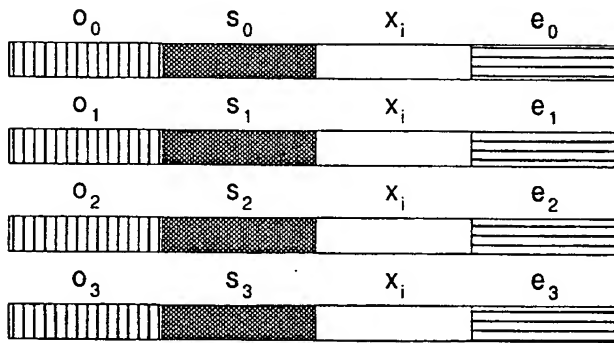


Abbildung 26

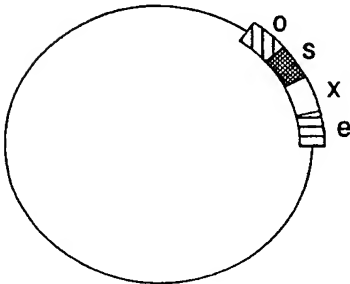


Abbildung 27

14 / 19

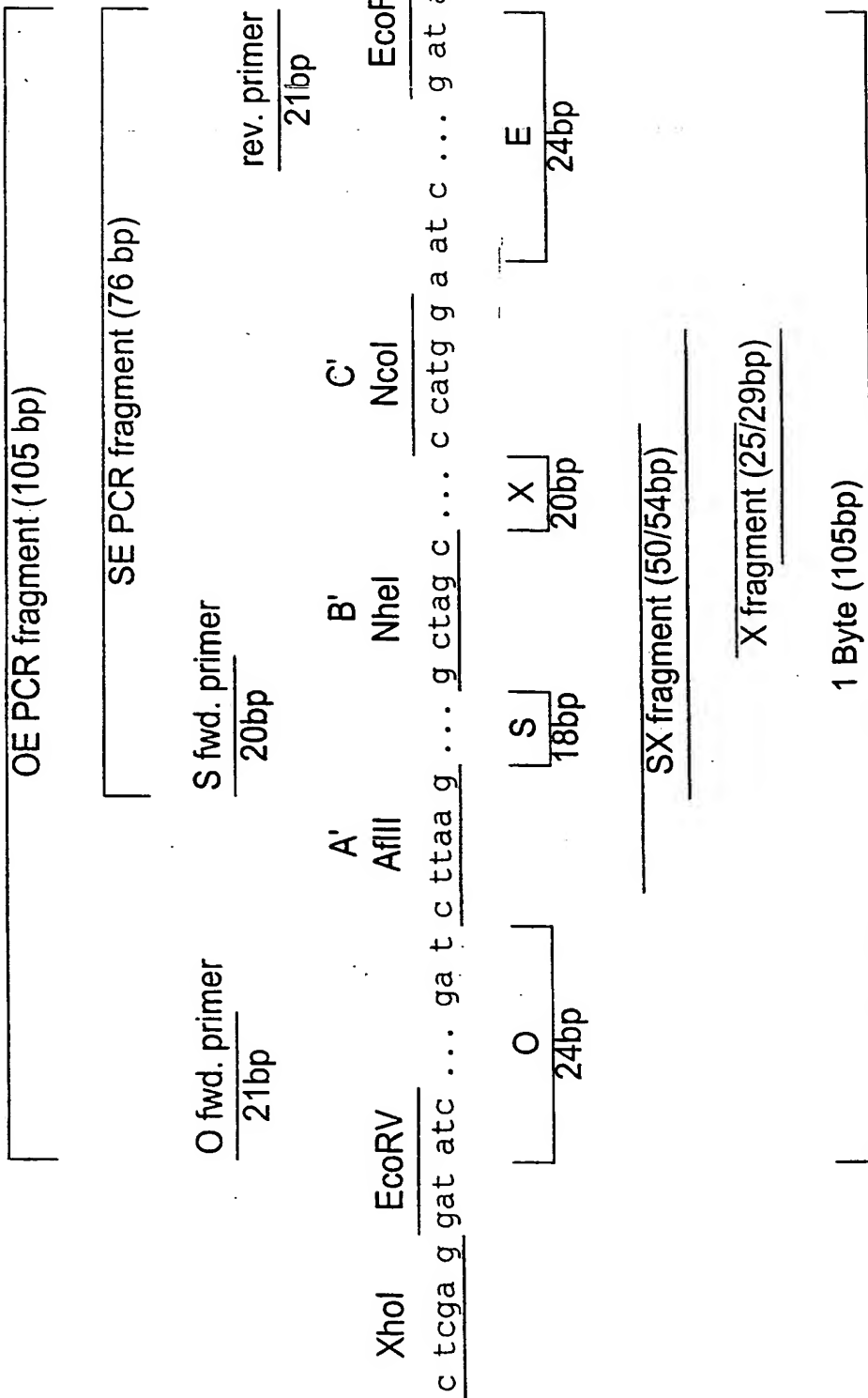


Abbildung 28

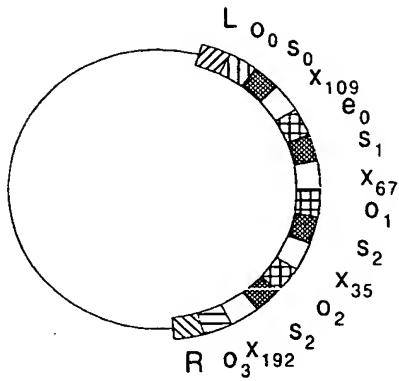


Abbildung 31

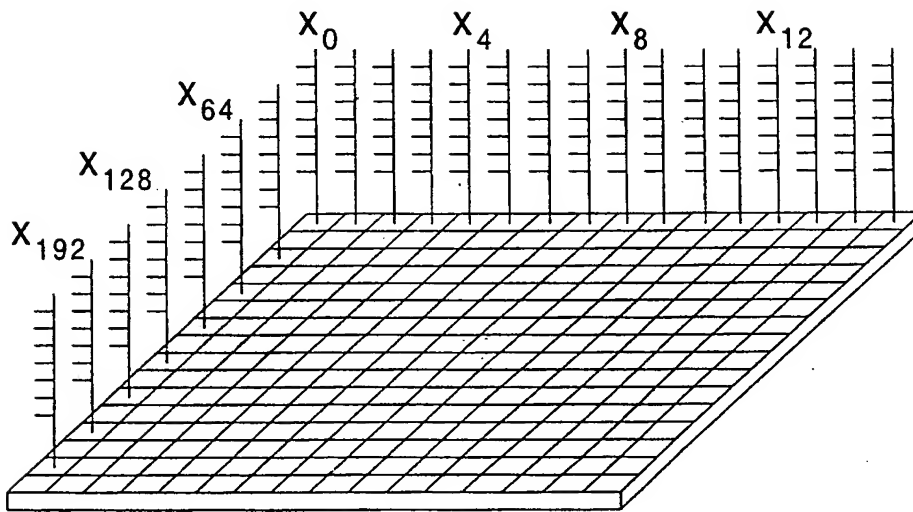


Abbildung 32

17 / 19

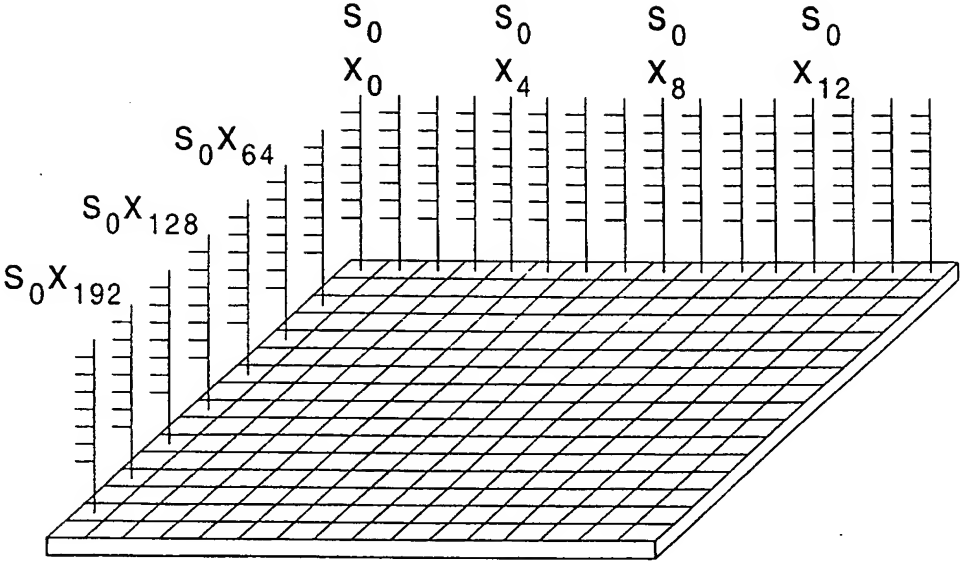


Abbildung 33

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Abbildung 34

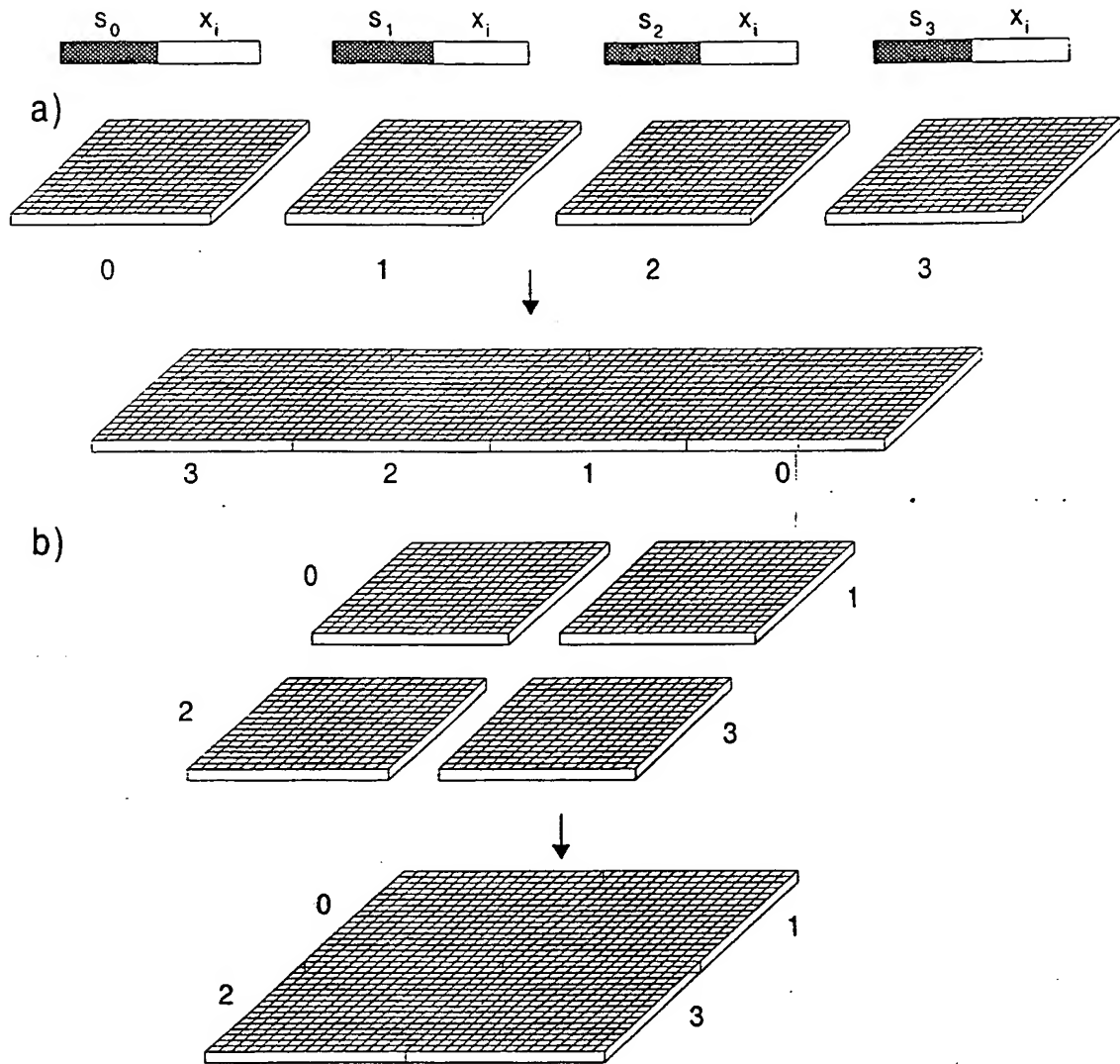


Abbildung 35

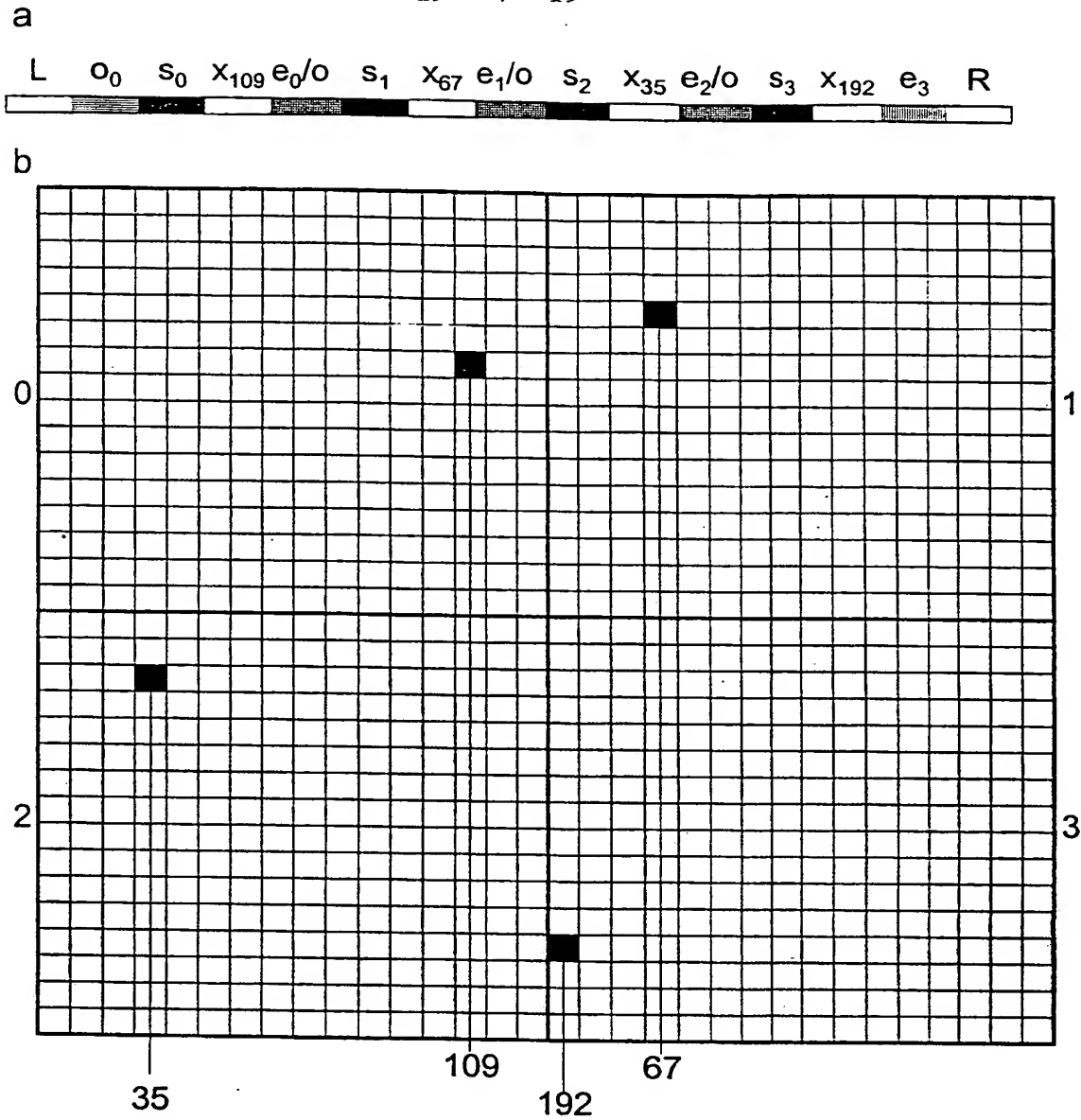


Abbildung 36

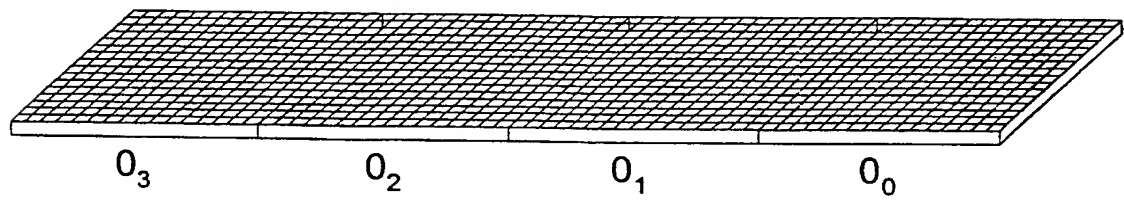


Abbildung 37

Sequenzprotokoll

Name	Sequenz
00	00U tcgaggatatccgggcttcgggtgaccgggatcttaaga
	00L agcttcttaagatcccgggtcaccggaagcccggatatcc
01	01U tcgaggatatcccttggcgggtgaggcggcgatcttaaga
	01L agcttcttaagatcgccgcctcaccgccaagggatatcc
02	02U tcgaggatatctgtgggctacggggccgcgatcttaaga
	02L agcttcttaagatcgcgccccgtagccacagatatcc
03	03U tcgaggatatctaggtggctgcgcggggcgatcttaaga
	03L agcttcttaagatcgccccgcgcagccacctagatatcc
S0	S0U ttaagcgaagtggcctggcgggctagcg
	S0L aattcgctagcgcccgcagggcacttcgc
S1	S1U ttaagaacccgagccccggacgggctagcg
	S1L aattcgctagccccgtccggggctcgggttc
S2	S2U ttaagcgcaccaggcagagcgcgctagcg
	S2L aattcgctagccgcgctctgcctggtgcgc
S3	S3U ttaagtcctcgtggctcgggcggcgctagcg
	S3L aattcgctagcgccgccccgaccacgaggac
E0	E0U gatcccatggaatcccttggcgggtgaggcggcgatatct
	E0L ctagagatatcgccgcctcaccgccaagggttccatggg
E1	E1U gatcccatggaatctgtgggctacggggccgcgatatct
	E1L ctagagatatcgcgccccgtagccacagattccatggg
E2	E2U gatcccatggaatctaggtggctgcgcggggcgatatct
	E2L ctagagatatcgccccgcgcagccacctagattccatggg
E3	E3U gatcccatggaatccgcaagccccaaggccccgatatct
	E3L ctagagatatcggggccttggggcttgcggttccatggg
X0	X0U ctagcgtatgccctgtagtttgagc
	X0L catggctccaaactacagggcatacg
X1	X1U ctagcgtaaagccaactgtaccctgc
	X1L catggcagggtagcagttggctttacg
X2	X2U ctagcgagactaagggtgttggtcgc
	X2L catggcgaccaacaaccttagtctcg
X3	X3U ctagcgttccttcaagtcactcgtcc
	X3L catgggacgagtgacttgaaggaacg
X4	X4U ctagccctgcccacagtagaataagc
	X4L catggcttattctactgtgggcaggg
X5	X5U ctagcgaatcaagactcgtgctaccc
	X5L catggggtagcacgagtccttgattcg
X6	X6U ctagcgatgacagttagacgcttccc
	X6L catggggaagcgtctaactgtcatcg
X7	X7U ctagcctcctgaacctaaagtcacgc
	X7L catggcgatgacttaggttcaggagg
X8	X8U ctagcccaaagttacagacctcagcc
	X8L catgggctgaggtctgtaactttggg

ERSATZBLATT (REGEL 26)

X9	X9U	ctagcgacctatgcaatctactcgcc
	X9L	catgggcgagtagattgcataggtcg
X10	X10U	ctagccagtagtaatcttccgcaggc
	X10L	catggcctgcggaagattactactgg
X11	X11U	ctagcctgttggtgggtactcctaac
	X11L	catggttaggagtacccaccaacagg
X12	X12U	ctagcaacgaggtggtacgacatagc
	X12L	catggctatgtcgtaccacctcgttg
X13	X13U	ctagccgcagtcattatagagagccc
	X13L	catggggctctctataatgactgcgg
X14	X14U	ctagcgtacttgtccctccagcatac
	X14L	catggtatgctggagggaagaagtacg
X15	X15U	ctagcatcctgtgtgcctaagtacgc
	X15L	catggcgtacttaggcacacaggatg
X16	X16U	ctagcgaggcacttgacgtatgactc
	X16L	catggagtcatacgtcaagtgcctcg
X17	X17U	ctagcgggatacctttctaagcgacc
	X17L	catgggtcgcttagaaaggtatcccg
X18	X18U	ctagcggataggtcgcagtcctgtac
	X18L	catggtacaagactgcgacctatccg
X19	X19U	ctagcgtacagtgcgaagagggtaac
	X19L	catggttaccctcttcgcactgtacg
X20	X20U	ctagcgcctccgacacagttactaac
	X20L	catggttagtaactgtgtcggaggcg
X21	X21U	ctagcgtgggtctggaacactattgc
	X21L	catggcaatagtgttccagaccacg
X22	X22U	ctagcaactcctcactgttcactgcc
	X22L	catgggcagtgaacagtgaggagttg
X23	X23U	ctagccgtatggtatatcgtcagccc
	X23L	catggggctgacgatataccatacgg
X24	X24U	ctagcgcacagaccaaggacattacc
	X24L	catgggtaatgtccttggtctgtgcg
X25	X25U	ctagcctcccactaaacggtcctatc
	X25L	catggataggaccgtttagtgggagg
X26	X26U	ctagccacggaccgagtatctattgc
	X26L	catggcaatagatactcgggccgtgg
X27	X27U	ctagcgtctgatgagcgcagttagtc
	X27L	catggactaactgcgctcatcagacg
X28	X28U	ctagcgagaatgcggtgatagtgacc
	X28L	catgggtcactatcacgcattctcg
X29	X29U	ctagcctacactaaccgacggaatgc
	X29L	catggcattccgtcggttagtgtagg
X30	X30U	ctagcaagtccttggtacgagtcgcc
	X30L	catgggggactcgtaacaaggacttg

X31	X31U	ctagcacctgaagttacgggaagtcc
	X31L	catgggacttcccgttaacttcaggtg
X32	X32U	ctagcaacagggttccttacagagggc
	X32L	catggccctctgtaaggaacctgttg
X33	X33U	ctagcacgtagccaagacttatcccc
	X33L	catgggggataagtcttggtactgtg
X34	X34U	ctagccagtcaggtctattgttgccc
	X34L	catggggcaacaatagacctgactgg
X35	X35U	ctagcgagaaccataatcgccctacc
	X35L	catgggtaggggcaatatgggtctcg
X36	X36U	ctagccataaaggaggagactgggctc
	X36L	catggagcccagtcctctttatgg
X37	X37U	ctagccatagtcctgcgaatactgccc
	X37L	catggggcagttatcgagactatgg
X38	X38U	ctagcgttagggcagccttaggtaac
	X38L	catggttacctaaggctgccctaacg
X39	X39U	ctagcatagatgtgacagtgggacgc
	X39L	catggcgtccactgtcacatctatg
X40	X40U	ctagcgattagagttgtaggcgctc
	X40L	catggagcgccctacaactctaactg
X41	X41U	ctagctggaagtacgtgtgggatagc
	X41L	catggctatcccacacgtacttcag
X42	X42U	ctagctgagatagtcgtggatgggtcc
	X42L	catgggaccatccacgactatctcag
X43	X43U	ctagcacgacctgttacggttacctc
	X43L	catggaggtaacggtaacaggtcgtg
X44	X44U	ctagccattgctggaactgctactcc
	X44L	catgggagtagcagttccagcaatgg
X45	X45U	ctagcgtacccccacccacataaagtc
	X45L	catggactttatgtgggtggggtacg
X46	X46U	ctagcagtgtaggaatacccgtgctc
	X46L	catggagcacgggtattcctacactg
X47	X47U	ctagccgttcgctatgttctacagcc
	X47L	catgggctgtagaacatagcgaacgg
X48	X48U	ctagcccctgtccccatacattactc
	X48L	catggagtaatgtatggggacagggg
X49	X49U	ctagcatgtaggcaactgtcgtcacc
	X49L	catgggtgacgacagttgcctacatg
X50	X50U	ctagcgtatgttcccagcgagtacac
	X50L	catgggtgtactcgctgggaacatacg
X51	X51U	ctagcgtgtaataccacgggtgttgcc
	X51L	catggcaacacccgtggtattacacg
X52	X52U	ctagctgtagtggtctttacgtggcc
	X52L	catgggccacgtaaagaccactacag

X53	X53U	ctagcgtcaggtgctcatccagttac
	X53L	catggtaactggatgagcacctgacg
X54	X54U	ctagcctccacaaacatacccacacc
	X54L	catgggtgtgggtatgtttgtggagg
X55	X55U	ctagctgaggtgtcaatgagagagggc
	X55L	catggcctctctcattgacacctcag
X56	X56U	ctagctcactcctctacgatgaacgc
	X56L	catggcggttcacgtagaggagttag
X57	X57U	ctagctcgtactgagatgttcctgcc
	X57L	catgggcaggaacatctcagtagag
X58	X58U	ctagctgactctggacttatgacgcc
	X58L	catgggcgtcataagtccagagtcag
X59	X59U	ctagcagattacggagtgtgttcccc
	X59L	catgggggaacacactccgtaatctg
X60	X60U	ctagccttaccactgttgagaggac
	X60L	catggtcctctgcaacagtggttaagg
X61	X61U	ctagcccgttatccgtactggaagtc
	X61L	catggacttccagtacggataacggg
X62	X62U	ctagcggttggtatcggtgtactc
	X62L	catggagtacagccgataacccaaacg
X63	X63U	ctagcgaggatgaacacaaggaggtc
	X63L	catggacctccttgtgttcacctcg
X64	X64U	ctagccatagcggatagggttacgtc
	X64L	catggacgtaaccctatccgctatgg
X65	X65U	ctagctctcatatccctaacctggcc
	X65L	catgggccaggttagggatatgagag
X66	X66U	ctagcgggtattgtctcgatgcacc
	X66L	catgggtgcatgacgagacaataaccg
X67	X67U	ctagcatcacctacctgagattgccc
	X67L	catggggcaatctcaggtaggtgatg
X68	X68U	ctagcgactggggtctgtgtttctc
	X68L	catggagaaacaacagaccccagtcg
X69	X69U	ctagcaggggtcggagttagagaatc
	X69L	catggattctctaactccgaccttg
X70	X70U	ctagcaccgaaggatgttgagtctcc
	X70L	catgggagactcaacatccttcggtg
X71	X71U	ctagcgggtgatgactcggaacttctc
	X71L	catggagaagttccgagtcacaccg
X72	X72U	ctagcgggtacacaaatggaaggtcc
	X72L	catgggaccttccatttgtgtacccg
X73	X73U	ctagcatcatctcgggtctacttcgc
	X73L	catggcgaagtagacccgagatgatg
X74	X74U	ctagcaatcagtagacatcgccctcc
	X74L	catgggagggcgatgtctactgattg

X75	X75U	ctagccacggaagtctattcacacgc
	X75L	catggcgtgtgaatagacttccgtgg
X76	X76U	ctagcaggggctgttacatgaagagc
	X76L	catggctcttcatgtaacagcccctg
X77	X77U	ctagccgtgacctgctgtgatttacc
	X77L	catgggtaaatcacagcaggtcacgg
X78	X78U	ctagcatagtcttagtccggcatcgc
	X78L	catggcgatgccggactaagactatg
X79	X79U	ctagcatcactacctgcctattcgcc
	X79L	catgggcgaataggcaggtagtgatg
X80	X80U	ctagcttaccttgtctaaccctcgcc
	X80L	catgggcgagggttagacaaggtaag
X81	X81U	ctagctcttaatccctctggctgacc
	X81L	catgggtcagccagagggattaagag
X82	X82U	ctagcagaccatccgagtaccatac
	X82L	catgggtatggtagctcggtgggtctg
X83	X83U	ctagcagactgaaccctcaagatgcc
	X83L	catgggcaccttgagggttcagtctg
X84	X84U	ctagcaggcattctgatactcctcgc
	X84L	catggcgaggagtatcagaatgcctg
X85	X85U	ctagcctgcaagagtcatatcacggc
	X85L	catggccgtgatatgactcttgacgg
X86	X86U	ctagcgtgggatttcgtctataccgc
	X86L	catggcggtatagacgaaatcccacg
X87	X87U	ctagccagaccgcctcctaattcttac
	X87L	catggtaagattaggaggcggctctgg
X88	X88U	ctagcatctccttgcttgacctcgc
	X88L	catggcgaggtacaagcaaggagatg
X89	X89U	ctagctgtgcggctatgtagtcctac
	X89L	catggtaggactacatagccgcacag
X90	X90U	ctagcagtgctcctaagcaaagacggc
	X90L	catggccgtctttgcttaggacactg
X91	X91U	ctagcttcccttctaccctgtaccac
	X91L	catgggtggtacagggtagaagggaag
X92	X92U	ctagcaagggtgtccagtggtcagatc
	X92L	catggatctgaccactggacaccttg
X93	X93U	ctagcatcagcatagggaccacactc
	X93L	catggagtgtggccctatgctgatg
X94	X94U	ctagcggacggccaactatcataagc
	X94L	catggcttatgatagttggccgtccg
X95	X95U	ctagctgaacctctgtgctgtaggac
	X95L	catggtcctacagcacagaggttcag
X96	X96U	ctagctgacttcacctggctgtctac
	X96L	catggtagacagccaggtgaagtcag

X97	X97U	ctagctcctaactctattgccgagcc
	X97L	catgggctcggcaatagagttaggag
X98	X98U	ctagccttatacctacgcgagcagatc
	X98L	catggatctgctcgcgtaggataagg
X99	X99U	ctagccgcataagcacatctgtagcc
	X99L	catgggctacagatgtgcttatgcgg
X100	X100U	ctagcaacgtagagaacgaagccagc
	X100L	catggctggcttcggttctctacgttg
X101	X101U	ctagctcacggagtcctgatgtatgc
	X101L	catggcatacatcaggactccgtgag
X102	X102U	ctagcatgtatcccgttggtgtaggc
	X102L	catggcctacaccaacgggatacatg
X103	X103U	ctagcgcctgactattttgacgggcc
	X103L	catgggaccgtcaaaatagtcaggcg
X104	X104U	ctagcgatagcacttcaaacgtcccc
	X104L	catgggggacgtttgaagtgtatcg
X105	X105U	ctagcgaagtgcattcttaccgtccc
	X105L	catggggacggtaagaatgcacttcg
X106	X106U	ctagcccttgctggttctaaactgggc
	X106L	catggcccagtttagaacgacaaggg
X107	X107U	ctagcgtttaaggagtgttcgtccgc
	X107L	catggcggacgaacactccttaaacg
X108	X108U	ctagccagagttaaagtggccccacc
	X108L	catgggtggggcaactttaactctgg
X109	X109U	ctagcgtcagaatcaatctcctggcc
	X109L	catggggccaggagattgattctgacg
X110	X110U	ctagcgtgcttgacctgaatccttcc
	X110L	catgggaaggattcaggtcaagcacg
X111	X111U	ctagcaccacccttcttcaccctatc
	X111L	catggatagggtgaagaagggtggtg
X112	X112U	ctagccagaaagggcagtcctaagtc
	X112L	catggcattaggactgccctttctgg
X113	X113U	ctagcaaccactatcgtagatccgc
	X113L	catggcggatgtacgatagtggttg
X114	X114U	ctagcgaataacttgggagttgcgagc
	X114L	catggctcgcaactcccaagtattcg
X115	X115U	ctagcggcaataagttcaggctgtcc
	X115L	catgggacagcctgaacttattgccg
X116	X116U	ctagcaaggagcagacagtcacgtc
	X116L	catggacatgactgtctgctcccttg
X117	X117U	ctagccataacacacgggacaacacc
	X117L	catgggtgttgctccgtgtgttatgg
X118	X118U	ctagccagaaacgaaggtcctaacgc
	X118L	catggcgtaggaccttcgtttctgg

X119	X119U	ctagcgtgcccaagccttaatagtgcc
	X119L	catgggcactattaaggcttggcacg
X120	X120U	ctagccggatgaggatgtatcaggtc
	X120L	catggacctgatacatcctcatccgg
X121	X121U	ctagccgtggcggttaaactcttaggc
	X121L	catggcctaagagtttaacgccacgg
X122	X122U	ctagcctgctcttgatggataaggcc
	X122L	catgggccttatccatcaagagcagg
X123	X123U	ctagcatcagagtggagagcacgatc
	X123L	catggatcgtgctctccactctgatg
X124	X124U	ctagcgcacaggaaatagaagtcgcc
	X124L	catgggcgacttctatttccctgtgcg
X125	X125U	ctagcccacccgactacagaacaatc
	X125L	catggattgttctgtagtcgggtggg
X126	X126U	ctagccgagattaccagttgtggtc
	X126L	catggaccacaactgggtaatctcgg
X127	X127U	ctagccgagttaatagacgcggaagc
	X127L	catggcttccgcgtctattaactcgg
X128	X128U	ctagccgctgcggtttctctattacc
	X128L	catgggtaatagagaaaccgcagcgg
X129	X129U	ctagctcagccgtaggtcttcaactc
	X129L	catggagttgaagacctacggctgag
X130	X130U	ctagcttagtggagcctcttcacgtc
	X130L	catggacgtgaagaggctccactaag
X131	X131U	ctagccagagtgtcggccttttaagc
	X131L	catggcttaaaaggccgacactctgg
X132	X132U	ctagcctaaaacttgactcgcgacc
	X132L	catgggtccgcgagtcgaagttttagg
X133	X133U	ctagcccgtctggcggataagatac
	X133L	catgggtatcttatccgaccagacggg
X134	X134U	ctagcagattggtcacaaactccaggc
	X134L	catggcctggagttgtgaccaatctg
X135	X135U	ctagcaccagaaggttggagaaggtc
	X135L	catggaccttctccaaccttctggtg
X136	X136U	ctagccacatcttacaccgtcatcgc
	X136L	catggcgatgacgggtgaagatgtgg
X137	X137U	ctagcgcgtcgggacttacaagatac
	X137L	catgggtatcttgtaagtcccgcgcg
X138	X138U	ctagcaggatggtgcaagcatactcc
	X138L	catgggagtatgcttgaccatcctg
X139	X139U	ctagcggacaaactgctggttcctac
	X139L	catggtaggaaccagcagtttgcgg
X140	X140U	ctagcgggtgcctcaaactggagtac
	X140L	catgggtactccagtttgaggcaaccg

X141	X141U	ctagctttgtaagacaccccagagcc
	X141L	catgggctctggggtgtcttcaaaag
X142	X142U	ctagcccccttgggtcgtagtgctac
	X142L	catggtagcactaacgaccaaagggg
X143	X143U	ctagcagagggctgagggatagaaac
	X143L	catggtttctatccctcagccctctg
X144	X144U	ctagcagttcgtacctttgacacgcc
	X144L	catgggcgtgtcaaaggtagcaactg
X145	X145U	ctagcgatttgtagtgataccacccc
	X145L	catgggggtggtatcactgcaaactcg
X146	X146U	ctagcgagtgtagaatccggctgaac
	X146L	catggttcagccggattctacactcg
X147	X147U	ctagcgcaagtccgtgttcctcttac
	X147L	catggtaagaggaacacggacttgcg
X148	X148U	ctagcctttgtacgggtggtgaaaccc
	X148L	catgggggtttcaccaccgtacaaagg
X149	X149U	ctagcactttcatgtggagacgctcc
	X149L	catgggagcgtctccacatgaaagtg
X150	X150U	ctagcaacactaagcggatgtcagcc
	X150L	catgggctgacatccgcttagtggtg
X151	X151U	ctagcctaactcaactccgcaacgtc
	X151L	catggacgttgcgaggttaggttagg
X152	X152U	ctagcaagtagatggaaaggggtgcc
	X152L	catggggcaccctttccatctacttg
X153	X153U	ctagcgaaatacagaggaacgagggtc
	X153L	catggaccctcgttcctcgtatttcg
X154	X154U	ctagcgaggctgctaagggaagaaagc
	X154L	catggctttttcccttagcagcctcg
X155	X155U	ctagcggacgttatttcacacctcgc
	X155L	catggcgaggtgtgaaataacgtccg
X156	X156U	ctagcgacacaagagcctaagccaac
	X156L	catggttggcttaggctcttgtgtcg
X157	X157U	ctagcgtgtttgtttaccctgccagc
	X157L	catggctggcagggtaaacaaacacg
X158	X158U	ctagcactgcgtcctcaciaaagaagc
	X158L	catggcttctttgtgaggacgcagtg
X159	X159U	ctagcagtccccgaagtcaaatagcc
	X159L	catgggctatttgacttcggggactg
X160	X160U	ctagcgatacttcccaatgcagacgc
	X160L	catggcgtctgcattgggaagtatcg
X161	X161U	ctagcggaactaatgtcatgctgccc
	X161L	catggggcagcatgacattagttccg
X162	X162U	ctagcgcatgggtgataacgtaagcc
	X162L	catgggcttacgttatcaccatcgcg

X163	X163U	ctagcctcctgtacggcaaatgttcc
	X163L	catgggaacatttgccgtacaggagg
X164	X164U	ctagcctatttggaaccgcacaagacc
	X164L	catgggtccttgtgcgggtccaaatagg
X165	X165U	ctagcatccttgtagcggacacgtac
	X165L	catggtacgtgtccgctacaaggatg
X166	X166U	ctagcaagtgggaataggggtccgtac
	X166L	catggtacggaccctattcccacttg
X167	X167U	ctagcgatagaaccgactgcaatgcc
	X167L	catgggcattgcagtcggttctatcg
X168	X168U	ctagcgcgaatgaagtggactatgcc
	X168L	catgggcatagtccacttcattcgcg
X169	X169U	ctagccgaactttagcatccgtgacc
	X169L	catgggtcacggatgctaaagttcgg
X170	X170U	ctagcgttagaaatgtctgcggtcgc
	X170L	catggcgaccgcagacatttctaacg
X171	X171U	ctagcgcaccacacgaaagaatctcc
	X171L	catgggagattctttcgtgtggtgcg
X172	X172U	ctagctcacactcttggtcggtatc
	X172L	catggatagccgaccaagagtgtgag
X173	X173U	ctagctcactctggccgaactgtatc
	X173L	catggatacagttcggccagagtgcg
X174	X174U	ctagctggagaacgggtcttgtctgtc
	X174L	catggacagacaagaccgttctccag
X175	X175U	ctagcaggtatcgcgtccctaagtgc
	X175L	catggacattagggacgcgatacctg
X176	X176U	ctagctgggtcactacacgaacatgc
	X176L	catggcatgttcgtgtagtgagccag
X177	X177U	ctagcgtaccagtgaatcgaggacac
	X177L	catgggtgtcccgatttcaactggtacg
X178	X178U	ctagccgaagaacaacatcagagccc
	X178L	catgggggtctgatgttcttctcgg
X179	X179U	ctagcgggtccaatcctaagcaagc
	X179L	catggcttgcttttaggattggacccg
X180	X180U	ctagcttgagcagaagaactctcgcc
	X180L	catgggagagagttcttctgctcaag
X181	X181U	ctagcattgccgtggtgtaactacgc
	X181L	catggcgtagttacaccacggcaatg
X182	X182U	ctagccgcatacgttctgtgcagtac
	X182L	catggtactgcacagaacgtatgcgg
X183	X183U	ctagccagcgatatggtggtgctatc
	X183L	catggatagcaccaccatctcgctgg
X184	X184U	ctagccgtaaacaccagagattggac
	X184L	catgggtccaatctctgggtgttacgg

X185	X185U	ctagctccacgataactcacaagggc
	X185L	catggcccttgtaggttatcgtggag
X186	X186U	ctagctcacttccttggtcgtaatgc
	X186L	catggcattacgaccaggggaagtgag
X187	X187U	ctagcgcggttggtcacctaatacggac
	X187L	catgggtccgtattaggtgacaacgcg
X188	X188U	ctagcatgttacactcgcacccgacc
	X188L	catgggtcggatgagagtgtaacatg
X189	X189U	ctagcatgtagccagtgattgtgccc
	X189L	catggggcacaatcactggctacatg
X190	X190U	ctagcaagatgcagagtctaccgcac
	X190L	catgggtgcggtagactctgcatcttg
X191	X191U	ctagcagaccgctctatcacaagcac
	X191L	catgggtgcttgtagatagagcggctg
X192	X192U	ctagcagggaggtccagaatcttcac
	X192L	catgggtgaagattctggacctccctg
X193	X193U	ctagcttagaccagcactgcgaagtc
	X193L	catggacttcgcagtgctggtctaag
X194	X194U	ctagcccagggcgtgactatacaaac
	X194L	catgggtttgtatagtcacgcctcggg
X195	X195U	ctagcgacaacctacaaacgctccac
	X195L	catgggtggagcgtttgtaggttgctg
X196	X196U	ctagccgtcattcgtcagaccagatc
	X196L	catggatctggtctgacgaatgacgg
X197	X197U	ctagcagatgagtgacgatggttgcc
	X197L	catgggcaaccatcgctcactcatctg
X198	X198U	ctagcaccacaaaccaggaaaagtcc
	X198L	catgggacttttcctggttggtgggtg
X199	X199U	ctagctgaggcatctttggagtacgc
	X199L	catggcgctactccaaagatgcctcag
X200	X200U	ctagctgtaatcggctcaggaagc
	X200L	catggcttgctgagaccgattacag
X201	X201U	ctagcctacagtcggattggctcaac
	X201L	catgggttgagccaatccgactgtagg
X202	X202U	ctagccttcaaggacctcgtgcatac
	X202L	catgggtatgcacgaggtccttgaagg
X203	X203U	ctagcggcaggataaagtgctgacac
	X203L	catgggtgtcagcactttatcctgccg
X204	X204U	ctagcgctggacgggttaaagggttac
	X204L	catggtaacctttaaccgtccaggcg
X205	X205U	ctagctgtccaagtaccaaagagcgc
	X205L	catggcgctctttggtacttgacag
X206	X206U	ctagctgcacactgggtctaacacac
	X206L	catgggtgtgtagaccagtggtgcag

X207 X207U ctagcaatctgtcgccctgcatagtgc
X207L catggcactatgcaggcgacagattg
X208 X208U ctagcccccttcgctcttctttcatcc
X208L catgggatgaaagaagagcgaagggg
X209 X209U ctagcaagatattctcccacgaccgc
X209L catggcggtcgtgggagaatatcttg
X210 X210U ctagcggtttctgcactacagccaac
X210L catggttggtgtagtgcagaaaccg
X211 X211U ctagcaaaggcatcggacatcctacc
X211L catgggtaggatgtccgatgcctttg
X212 X212U ctagcacgaagtatgcgttgtgagcc
X212L catgggctcacaacgcatacttcgtg
X213 X213U ctagcggacatcggatttgataccc
X213L catggggtatccaaatccgatgtccg
X214 X214U ctagccggtacttttgcgactctgac
X214L catggtcagagtcgcaaaagtaccgg
X215 X215U ctagcggttacgctcaccgaaatctc
X215L catggagatttcggtgagcggtaacg
X216 X216U ctagcaagaagacgatgtgtgctcgc
X216L catggcgagcacacatcgtcttcttg
X217 X217U ctagcggctattcattctcgcctac
X217L catggtaggcgagaatgaatagccg
X218 X218U ctagcaagggaatgtcagttctgccc
X218L catggggcagaactgacattcccttg
X219 X219U ctagcggcttgtaggcccgttaagac
X219L catggtcttaacggcctaacaagccg
X220 X220U ctagcaaggatagagatgccgttgcc
X220L catgggcaacggcatctctatccttg
X221 X221U ctagcagccccgacatgacagatacac
X221L catggtgtatctgtcatgtcgggctg
X222 X222U ctagcttgagagactttgctgacgc
X222L catggcgtcagcaaagtctctccaag
X223 X223U ctagctgcgtgtacgggcttatagac
X223L catggtctataagcccgtacacgcag
X224 X224U ctagctcttattcgttagccgaggcc
X224L catgggcctcggctaacgaataagag
X225 X225U ctagcagccgaacaggacaaaactcc
X225L catgggagttttgtcctgttcggctg
X226 X226U ctagccatcattaacttcccctggcc
X226L catgggccaggggaagttaatgatgg
X227 X227U ctagcccgaactaaatgtcacagcgc
X227L catggcgctgtgacatttagttcggg
X228 X228U ctagcgacgcaatcttaaccaacccc
X228L catgggggttggttaagattgcgtcg

X229	X229U	ctagctgtatgcacatcaccgagtg
	X229L	catggcactcggatgatgtgcatacag
X230	X230U	ctagccgtctcattctggtcattgcc
	X230L	catgggcaatgaccagaatgagacgg
X231	X231U	ctagcaggttgggttcccctatcatc
	X231L	catggatgataggggaacccaacctg
X232	X232U	ctagccggttgcttatcaggaatccc
	X232L	catggggattcctgataagcaaccgg
X233	X233U	ctagcacggccagtaagtgtccaatc
	X233L	catggattggacacttactggccgtg
X234	X234U	ctagcacacctgggcgtgaataactc
	X234L	catggagttattcacgcccaggtgtg
X235	X235U	ctagccttgggctcctgctaaaagac
	X235L	catggtcttttagcaggagcccaagg
X236	X236U	ctagcgcgacctctaagccttgaaac
	X236L	catggtttcaaggcttagaggtcgcg
X237	X237U	ctagccaggtttgggggtacatgatc
	X237L	catggatcatgtacccccaaacctgg
X238	X238U	ctagctcgttgtctgccctcttgtag
	X238L	catggtacaagagggcagacaacgag
X239	X239U	ctagcagacctcggttaaccacgaaac
	X239L	catggtttcgtgggttaccgaggtctg
X240	X240U	ctagcagtaagggtctttgggtgccac
	X240L	catggtggcacccaaagacctactg
X241	X241U	ctagcagtagccttgtgggaaccaac
	X241L	catggttgggtcccacaaggctactg
X242	X242U	ctagcgattaacctctcggaatgcgc
	X242L	catggcgcattccgagaggttaatcg
X243	X243U	ctagcgcgatattgttccctgcttcc
	X243L	catgggaagcaggggaacaatatcgcg
X244	X244U	ctagccaaccgtgttagcgaccatac
	X244L	catggtatggtcgctaacacggttgg
X245	X245U	ctagcgaatgggtggtggaaggatac
	X245L	catggtatccttccaccaccattcg
X246	X246U	ctagcaccaagtcgctgtcaactgac
	X246L	catggtcagttgacagcgacttggtg
X247	X247U	ctagccattttcaggaaggagacggc
	X247L	catggccgtctccttccctgaaaatgg
X248	X248U	ctagctgggttgactgatacgaacgc
	X248L	catggcggttcgtatcagtccaaccag
X249	X249U	ctagcattcacccaatggtctacggc
	X249L	catggccgtagaccattgggtgaatg
X250	X250U	ctagcgggtggtaggcataatccgaaac
	X250L	catggtttcggatatgcctaccaccg

X251	X251U	ctagctgtctgtgaccttgggattgc
	X251L	catggcaatcccaaggtcacagacag
X252	X252U	ctagccaagcctcaatgcctaaaggc
	X252L	catggccttttaggcattgaggcttgg
X253	X253U	ctagcacttcttacgtcaccgcgatc
	X253L	catggatcgcggtgacgtaagaagtg
X254	X254U	ctagcggttcctcaccattcggttac
	X254L	catggtaaccgaatggtgaggaaccg
X255	X255U	ctagcagagtcaatgtcggttggctc
	X255L	catggagccaaccgacattgactctg

Name	Algo mer
O0	tcgaggatatccgggcttcgggtgaccgggatcttaaga cctataggcccggaagcccactggccctagaattcttcga
O1	tcgaggatatcccttggcgggtgaggcggcgatcttaaga cctatagggaaccgcccactccgcccctagaattcttcga
O2	tcgaggatatctgtgggctacggggccgcgatcttaaga cctatagacacccgatgccccggcgctagaattcttcga
O3	tcgaggatatctaggtggctgcgcgggcgatcttaaga cctatagatccaccgacgcgccccgctagaattcttcga
S0	ttaagcgaagtggcctggcgggcgctagcg cgcttcaccggaccgcccgcgatcgcttaa
S1	ttaagaacccgagccccggacgggctagcg cttgggctcggggcctgcccgatcgcttaa
S2	ttaagcgcaccaggcagagcgcggtagcg cgcggtggtccgtctcgcgccgatcgcttaa
S3	ttaagtccctcgtggtcgggcggcgatagcg caggagcaccagcccgcgcgatcgcttaa
E0	gatccccatggaatcccttggcgggtgaggcggcgatatct gggtaccttagggaaccgccactccgcccctatagagatc
E1	gatccccatggaatctgtgggctacggggccgcgatatct gggtaccttagacacccgatgccccggcgctatagagatc
E2	gatccccatggaatctaggtggctgcgcgggcgatctct gggtaccttagatccaccgacgcgccccgctatagagatc
E3	gatccccatggaatccgcaagccccaaggccccgatatct gggtaccttaggcgttcgggggtccggggctatagagatc
X0	ctagcgtatgccctgtagtttgagc gcatacgggacatcaaacctcggtac
X1	ctagcgtaaagccaactgtaccctgc gcatttcgggttgacatgggacgggtac
X2	ctagcgagactaaggttggttggtcgc gctctgattccaacaaccagcggtac
X3	ctagcgttccttcaagtcactcgtec gcaaggaaagttcagtgagcagggtac
X4	ctagccctgcccacagtagaataagc gggacgggtgtcatcttattcggtac
X5	ctagcgaatcaagactcgtgctaccc gcttagttctgagcacgatgggggtac
X6	ctagcgatgacagttagacgcttccc gctactgtcaatctgcgaaggggtac
X7	ctagccctcctgaacctaaagtcacgc ggaggacttgattcagtagcggtac
X8	ctagcccaaagttacagacctcagcc gggtttcaatgtctggagtcgggtac

X9 ctagcgacctatgcaatctactcgcc
 gctggatacgttagatgagcgggtac
X10 ctagccagtagtaatcttccgcaggc
 ggtcatcattagaaggcgtccggtac
X11 ctagcctggtggtgggtactcctaac
 ggacaaccacccatgaggattggtac
X12 ctagcaacgaggtggtacgacatagc
 gttgctccaccatgctgtatcggtac
X13 ctagccqcgatcattatagagagccc
 ggcgtcagtaatatctctcgggggtac
X14 ctagcgtacttgctccctccagcatac
 gcatgaacagggaggtcgtatggtac
X15 ctagcatcctgtgtgcctaagtacgc
 gtaggacacacggattcatgcggtac
X16 ctagcgaggcacttgacgtatgactc
 gctccgtgaactgcatactgagggtac
X17 ctagcgggatacctttctaagcgacc
 gccctatggaaagattcgctgggtac
X18 ctagcggataggtcgcagtccttgtagc
 gcctatccagcgtcagaacatggtac
X19 ctagcgtacagtgccaagagggtaac
 gcatgtcacgcttctcccattggtac
X20 ctagcgctccgacacagttactaac
 gcgagggtgtgtcaatgattggtac
X21 ctagcgtgggtctggaacactattgc
 gcacccagaccttggtgataacggtac
X22 ctagcaactcctcactgttcactgcc
 gttgaggagtgacaagtgcggtac
X23 ctagccgtatggtatatcgtcagccc
 ggcataccatatagcagtcgggggtac
X24 ctagcgcacagaccaaggacattacc
 gcgtgtctggttcctgtaatgggtac
X25 ctagcctcccactaaacgggtcctatc
 ggaggggtgatttgccaggataggtac
X26 ctagccacggaccgagtatctattgc
 ggtgcctggctcatagataacggtac
X27 ctagcgtctgatgagcgcagttagtc
 gcagactactcgcgtcaatcagggtac
X28 ctagcgagaatgcggtgatagtgacc
 gctcttacgccactatcactgggtac
X29 ctagcctacactaaccgacggaatgc
 ggatgtgattggctgccttacggtac
X30 ctagcaagtccttggttacgagtcccc
 gttcaggaacaatgctcaggggggtac

16

X31 ctagcacctgaagttacgggaagtc
gtggacttcaatgcccttcagggtac
X32 ctagcaacaggttccttacagagggc
gttgtccaaggaatgtctcccgggtac
X33 ctagcacgtagccaagacttatcccc
gtgcatcgggttctgaataggggggtac
X34 ctagccagtcaggtctattgttgccc
ggtcagtcagataacaacgggggtac
X35 ctagcgagaaccatattgcccctacc
gctcttgggtataacggggatgggtac
X36 ctagccataaagaggagactgggctc
ggtatttctcctctgaccgaggtac
X37 ctagccatagtctgcgaatactgccc
ggtatcagacgcttatgacgggggtac
X38 ctagcgttagggcagccttaggtaac
gcaatcccgtcgggaatccattgggtac
X39 ctagcatagatgtgacagtgggacgc
gtatctacactgtcaccctgcgggtac
X40 ctagcgattagagttgtagggcgctc
gctaattctcaacatcccgcgaggtac
X41 ctagctggaagtacgtgtgggtagc
gaccttcatgcacaccctatcggtac
X42 ctagctgagatagtcgtggatgggtcc
gactctatcagcacctaccagggtac
X43 ctagcacgacctgttaccgttacctc
gtgctggacaatggcaatggaggtac
X44 ctagccattgctggaactgctactcc
ggtaacgaccttgacgatgagggtac
X45 ctagcgtacccccacccacataaagtc
gcatgggggtgggtgtatttcagggtac
X46 ctagcagtgtaggaatacccggtc
gtcacatccttatgggcacgaggtac
X47 ctagccgttcgctatgttctacagcc
ggcaagcgatacaagatgtcgggtac
X48 ctagcccctgtccccatacattactc
ggggacaggggtatgtaatgaggtac
X49 ctagcatgtaggcaactgtcgtcacc
gtacatccgttgacagcagtgagggtac
X50 ctagcgtatgttcccagcgagtacac
gcatacaaggggtcgtcatgtgggtac
X51 ctagcgtgtaataccacgggtgttgc
gcacattatgggtgcccacaacgggtac
X52 ctagctgtagtggtctttacgtggcc
gacatcaccagaaatgcaccgggtac

ERSATZBLATT (REGEL 26)

X53 ctagcgtcaggtgctcatccagttac
gcagtcacgagtaggtcaatggtac
X54 ctagcctccacaaacataccacacc
ggaggtgtttgtatgggtgtgggtac
X55 ctagctgaggtgtcaatgagagaggc
gactccacagttactctctccggtac
X56 ctagctcactcctctacgatgaacgc
gagtgaggagatgctacttgcggtac
X57 ctagctcgtactgagatgttcctgcc
gagcatgactctacaaggacgggtac
X58 ctagctgactctggacttatgacgcc
gactgagacctgaatactgcgggtac
X59 ctagcagattacggagtgtgttcccc
gtctaatagcctcacacaagggggtac
X60 ctagccttaccactgttgcagaggac
ggaatggtgacaacgtctcctggtac
X61 ctagcccgttatccgtactggaagtc
gggcaataggcatgaccttcaggtac
X62 ctagcgtttgggtatcggctgtactc
gcaaaccatagccgacatgaggtac
X63 ctagcgaggatgaacacaaggaggtc
gctcctacttgtgttcctccaggtac
X64 ctagccatagcggatagggttacgtc
ggtatcgccatcccaatgcaggtac
X65 ctagctctcatatccctaacctggcc
gagagtatagggttgaccgggtac
X66 ctagcggatttgtctcgtcatgcacc
gccataacagagcagtagctgggtac
X67 ctagcatcacctacctgagattgccc
gtagtggatggactctaacggggtac
X68 ctagcgactggggtctgttgtttctc
gctgacccagacaacaaagaggtac
X69 ctagcaggggtcggagttagagaatc
gtccccagcctcaatctcttaggtac
X70 ctagcaccgaaggatgttgagtctcc
gtggcttcctacaactcagagggtac
X71 ctagcgggtgatgactcggaacttctc
gccactactgagccttgaagaggtac
X72 ctagcgggtacacaaatggaagggtcc
gcccattgtgtttaccttccagggtac
X73 ctagcatcatctcgggtctacttcgc
gtagtagagcccagatgaagcgggtac
X74 ctagcaatcagtagacatcgccctcc
gttagtcactctgtagcgggagggtac

X75 ctagccacggaagtctattcacacgc
ggtgccttcagataagtgtgcggtac
X76 ctagcaggggctgttacatgaagagc
gtcccccacaatgtacttctcggtac
X77 ctagccgtgacctgctgtgatttacc
ggcactggacgacactaaatgggtac
X78 ctagcatagtcttagtccggcatcgc
gtatcagaatcaggccgtagcggtac
X79 ctagcatcactacctgcctattcgcc
gtagtgatggacggataagcggtac
X80 ctagcttaccttgtctaaccctcgcc
gaatggaacagattgggagcggtac
X81 ctagctcttaatccctctggctgacc
gagaattagggagaccgactgggtac
X82 ctagcagaccatccgagtaccatac
gtctgggtaggctcatggatgggtac
X83 ctagcagactgaaccctcaagatgcc
gtctgacttgggagttctacgggtac
X84 ctagcaggcattctgatactcctcgc
gtccgtaagactatgaggagcggtac
X85 ctagcctgcaagagtcatatcacggc
ggacgttctcagtatagtgccggtac
X86 ctagcgtgggatttcgtctataccgc
gcaccctaaagcagatatggcggtac
X87 ctagccagaccgcctcctaattctac
ggtctggcgaggattagaatgggtac
X88 ctagcatctccttgtgttacctcgc
gtagaggaacgaacatggagcggtac
X89 ctagctgtgcggctatgtagtcctac
gacacgccgatacatcaggatgggtac
X90 ctagcagtgtcctaagcaaagacggc
gtcacaggattcgtttctgccggtac
X91 ctagcttcccttctaccctgtaccac
gaagggaagatgggacatggtgggtac
X92 ctagcaagggtgtccagtggtcagatc
gtccacagggtcaccagtctaggtac
X93 ctagcatcagcatagggaccacactc
gtagtcgtatccctgggtgtgaggtac
X94 ctagcggacggccaactatcataagc
gcctgccggttgatagtattcggtac
X95 ctagctgaacctctgtgctgtaggac
gacttggagacacgacatcctgggtac
X96 ctagctgacttcacctggctgtctac
gactgaagtggaccgacagatgggtac

X97 ctagctcctaactctattgccgagcc
gaggattgagataacggctcgggtac
X98 ctagccttatcctacgcgagcagatc
ggaataggatgcgctcgtctaggtagc
X99 ctagccgcataagcacatctgtagcc
ggcgtattcgtgtagacatcgggtac
X100 ctagcaacgtagagaacgaagccagc
gttgcatctcttgccttcggtcgggtac
X101 ctagctcacggagtcctgatgtatgc
gagtgcctcaggactacatacgggtac
X102 ctagcatgtatcccgttggtgtaggc
gtacatagggaaccacatccgggtac
X103 ctagcgcctgactatcttgacgggtcc
gcggactgataaaactgccagggtac
X104 ctagcgatagcacttcaaacgtcccc
gctatcgtgaagtttgagggggtac
X105 ctagcgaagtgcattcttaccgtccc
gcttcacgtaagaatggcagggggtac
X106 ctagcccttgctcgttctaaactgggc
gggaacagcaagatttgaccgggtac
X107 ctagcgtttaaggagtgttcgtccgc
gcaaattcctcacaagcaggcgggtac
X108 ctagccagagttaaagttgccccacc
ggtctcaatttcaacgggggtgggtac
X109 ctagcgtcagaatcaatctcctggcc
gcagtcttagttagaggaccgggtac
X110 ctagcgtgcttgacctgaatccttcc
gcacgaactggacttaggaagggtac
X111 ctagcaccacccttcttcaccctatc
gtgggtgggaagaagtgggataggtagc
X112 ctagccagaaagggcagtcctaatgc
ggtctttcccgtcaggattacgggtac
X113 ctagcaaccactatcgtacatccgc
gttgggtgatagcatgtaggcgggtac
X114 ctagcgaatacttgggagttgcgagc
gcttatgaaccctcaacgctcgggtac
X115 ctagcggcaataagttcaggctgtcc
gccgttattcaagtcgacagggtac
X116 ctagcaaggagcagacagtcacgtc
gttcctcgtctgtcagtagcagggtac
X117 ctagccataacacacgggacaacacc
ggtattgtgtgcctgttggtgggtac
X118 ctagccagaaacgaagtcctaaccgc
ggtctttgcttccaggattgcgggtac

X119 ctagcgtgcccaagccttaatagtgcc
gcacgggttcggaattatcacgggtac
X120 ctagccggatgaggatgtatcaggtc
ggcctactcctacatagtccaggtag
X121 ctagccgtggcggttaaactcttaggc
ggcaccgcaatttgagaatccggtag
X122 ctagcctgctcttgatggataaggcc
ggacgagaactacattccgggtac
X123 ctagcatcagagtggagagcacgac
gtagtctcacctctcgtgctaggtag
X124 ctagcgcacaggaaatagaagtcgcc
gcgtgtcctttatcttcagcgggtac
X125 ctagcccacccgactacagaacaatc
gggtgggctgatgtcttgttaggtac
X126 ctagccgagattaccagttgtggtc
ggctctaattgggtcaacaccaggtag
X127 ctagccgagttaatagacgcggaagc
ggctcaattatctgcgccttcggtag
X128 ctagccgctgcggtttctctattacc
ggcgacgcaaagagataatgggtac
X129 ctagctcagccgtaggtcttcaactc
gagtcggcatccagaagttgaggtag
X130 ctagcttagtggagcctcttcacgtc
gaatcacctcggagaagtgcaggtag
X131 ctagccagagtgtcggccttttaagc
ggtctcacagccgaaaattcggtag
X132 ctagcctaaaacttgactcgcggacc
ggattttgaaactgagcgcctgggtac
X133 ctagcccgtctggtcggataagatac
gggcagaccagcctattctatggtag
X134 ctagcagattggtcacaactccaggc
gtctaaccagtggtgaggtccggtag
X135 ctagcaccagaaggttgagagaaggtc
gtggtcttccaacctcttcaggtag
X136 ctagccacatcttacaccgtcatcgc
ggtgtagaatgtggcagtagcggtag
X137 ctagcgcgtcgggacttacaagatac
gcgcagccctgaatgttctatggtag
X138 ctagcaggatgggtgcaagcatactcc
gtcctaccacgttcgtatgaggtag
X139 ctagcggacaaactgctggttcctac
gcctgtttgacgaccaaggatggtag
X140 ctagcggttgccctcaaactggagtag
gccaacggagtttgacctcatggtag

X141 ctagctttgtaagacaccccagagcc
gaaacattctgtggggtctcgggtac
X142 ctagccccctttggtcgttagtgtac
ggggaaaccagcaatcacgatggtac
X143 ctagcagagggctgaggatagaaac
gtctcccgaactccctatctttggtac
X144 ctagcagttcgtacctttgacacgcc
gtcaagcatggaaactgtgcgggtac
X145 ctagcgatttgcagtataccacccc
gctaaacgtcactatggtgggggtac
X146 ctagcgagtgtagaatccggctgaac
gtcacatcttaggccgacttggtac
X147 ctagcgcaagtccgtgttcctcttac
gcgttcaggcacaaggagaatggtac
X148 ctagcctttgtacggtggtgaaaccc
ggaaacatgccaccactttgggggtac
X149 ctagcactttcatgtggagacgtcc
gtgaaagtacacctctgcgagggtac
X150 ctagcaacactaagcggtatgtcagcc
gttggtgattcgcctacagtcgggtac
X151 ctagcctaactcaactccgcaacgtc
ggattgagttgaggcgttgaggtac
X152 ctagcaagtagatggaaagggtgccc
gttcactctacctttcccacgggggtac
X153 ctagcgaaatacagaggaacgagggtc
gctttatgctccttgcctccaggtac
X154 ctagcgaggctgctaagggaaaaagc
gtccgacgattccctttttcggtac
X155 ctagcggacgttatttcacacctcgc
gcctgcaataaagtgtggagcggtac
X156 ctagcgacacaagagcctaagccaac
gctgtgttctcggattcggttggtac
X157 ctagcgtgtttgtttaccctgccagc
gcacaaacaaatgggacggtcggtac
X158 ctagcactgcgtcctcacaagaagc
gtgacgcaggagtgtttcttcggtac
X159 ctagcagtcctccgaagtcaaatagcc
gtcaggggcttcagtttatcggtac
X160 ctagcgatacttcccaatgcagacgc
gctatgaagggttacgtctgcgggtac
X161 ctagcggaactaatgtcatgctgcc
gccttgattacgtacgacgggggtac
X162 ctagcgcgatggtgataacgtaagcc
gcgctaccactattgcattcggtac

X163 ctagcctcctgtacggcaaatgttcc
ggaggacatgccgtttacaagggtag
X164 ctagcctatTTggaccgcacaagacc
ggataaacctggcgtgttctgggtac
X165 ctagcatccttgtagcggacacgtac
gtaggaacatcgctgtgcatggtag
X166 ctagcaagtgggaatagggtccgtac
gttcacccttatcccaggcatggtag
X167 ctagcgatagaaccgactgcaatgcc
gctatcttggctgacgttacgggtac
X168 ctagcgcaatgaagtggactatgcc
gcgcttacttcacctgatacgggtac
X169 ctagccgaactttagcatccgtgacc
ggcttgaaatcgtaggcactgggtac
X170 ctagcgttagaaatgtctgcggtcgc
gcaatctttacagacgccagcgtac
X171 ctagcgaccacacgaaagaatctcc
gcgtggtgtgctttcttagagggtac
X172 ctagctcacactcttggctcggtatc
gagtgtgagaaccagccgataggtag
X173 ctagctcactctggccgaactgtatc
gagtgagaccggcttgacataggtag
X174 ctagctggagaacggtcttgtctgtc
gacctcttgccagaacagacaggtag
X175 ctagcaggtatcggtccctaagtgc
gtccatagcgcagggttacaggtac
X176 ctagctggctcactacacgaacatgc
gaccgagtgatgtgcttgtacggtag
X177 ctagcgtaccagtgaaatcgggacac
gcatggctacttttagccctgtggtag
X178 ctagccgaagaacaacatcagagccc
ggcttcttgttgtagtctcggggtag
X179 ctagcgggtccaatcctaaagcaagc
gcccagggttaggatttcgttcggtag
X180 ctagcttgagcagaagaactctcgcc
gaactcgtcttcttgagagcgggtac
X181 ctagcattgccgtggtgtaactacgc
gtaacggcaccacattgatgcggtag
X182 ctagccgcatacgttctgtgcagtac
ggcgtatgcaagacacgtcatggtag
X183 ctagccagcgatatggtggtgctatc
ggtcgctataccaccacgataggtag
X184 ctagccgtaacaccagagattggac
ggcattgtgggtctctaacctggtag

X185 ctagctccacgataactcacaagggc
gaggtgctattgagtgttcccggtag
X186 ctagctcacttccctggcgtaatgc
gagtgaagggaccagcattacggtag
X187 ctagcgcggttgacctaataacggac
gcgcaacagtggattatgcctggtag
X188 ctagcatgttacactcgcacccgacc
gtacaatgtgagcgtaggctgggtac
X189 ctagcatgtagccagtgattgtgccc
gtacatcggtcactaacacggggtag
X190 ctagcaagatgcagagtctaccgcac
gttctacgtctcagatggcgtggtag
X191 ctagcagaccgctctatcacaagcac
gtctggcgagatagtgttcgtggtag
X192 ctagcagggaggtccagaatcttcac
gtccctccaggtcttagaagtggtag
X193 ctagcttagaccagcactgcgaagtc
gaatctggtcgtgacgcttcaggtac
X194 ctagcccgaggcgtgactatacaaac
gggctccgcactgatatgtttggtag
X195 ctagcgacaacctacaaacgctccac
gctgttggtatgtttgcgaggtggtag
X196 ctagccgtcattcgtcagaccagatc
ggcagtaagcagtctggctaggtac
X197 ctagcagatgagtgcgatgggtgcc
gtctactcactgctaccaacgggtac
X198 ctagcaccacaaaccaggaaaagtc
gtgggtgttggtccttttcagggtac
X199 ctagctgaggcatctttggagtacgc
gactccgtagaaacctcatgcggtag
X200 ctagctgtaatcggctctcaggcaagc
gacattagccagagtcggttcggtag
X201 ctagcctacagtcggattggctcaac
ggatgtcagcctaaccgagttggtag
X202 ctagccttcaaggacctcgtgcatac
ggaagttcctggagcacgtatggtag
X203 ctagcggcaggataaagtgcagacac
gccgtcctatttcacgactgtggtag
X204 ctagcgccctggacggttaaagggtac
gcggacctgccaatttccaatggtag
X205 ctagctgtccaagtacaaagagcgc
gacaggttcagtgtttctcgcggtag
X206 ctagctgcacactgggtctaacacac
gacgtgtgaccagattgtgtggtag

X207 ctagcaatctgtcgccctgcatagtgc
gtagacagcggacgtatcacggtac
X208 ctagcccttcgctcttctttcatcc
ggggaagcgagaagaaagtagggtac
X209 ctagcaagatattctcccacgaccgc
gttctataagaggggtgctggcggtac
X210 ctagcggtttctgcactacagccaac
gccaaagacgtgatgtcggttggtac
X211 ctagcaaaggcatcggacatcctacc
gtttccgtagcctgtaggatgggtac
X212 ctagcacgaagtatgcgttgtagacc
gtgcttcatacgaacactcgggtac
X213 ctagcggacatcggatttggtaccc
gcctgtagcctaaccctatgggtac
X214 ctagccggtacttttgcgactctgac
ggccatgaaaacgctgagactggtac
X215 ctagcgttaccgctcaccgaaatctc
gcaatggcgagtggctttagaggtac
X216 ctagcaagaagacgatgtgtgctcgc
gttcttctgctacacacgagcggtac
X217 ctagcggctattcattctcgccctac
gccgataagtaagagcgggatgggtac
X218 ctagcaagggaatgtcagttctgccc
gttcccttacagtcaagacgggggtac
X219 ctagcggcttgtaggcggttaagac
gccgaacaatccggcaattctggtac
X220 ctagcaaggatagagatgccgttgcc
gttcctatctctacggcaacgggtac
X221 ctagcagcccgcacatgacagatacac
gtcgggctgtactgtctatgtggtac
X222 ctagcttgagagactttgctgacgc
gaacctctctgaaacgactgcggtac
X223 ctagctgcgtgtacgggcttatagac
gacgcacatgccgaatatctggtac
X224 ctagctcttattcgttagccgaggcc
gagaataagcaatcggtccgggtac
X225 ctagcagccgaacaggacaaaactcc
gtcggcttgctcctgttttgaggggtac
X226 ctagccatcattaacttcccctggcc
ggtagtaattgaaggggaccgggtac
X227 ctagcccgaactaaatgtcacagcgc
gggcttgatttacagtgtcgcggtac
X228 ctagcgacgcaatcttaaccaacccc
gctgcgttagaattgggtgggggtac

X229 ctagctgtatgcacatcaccgagtgc
gacatacgtgtagtggctcacggtac
X230 ctagccgtctcattctggtcattgcc
ggcagagtaagaccagtaacgggtac
X231 ctagcaggttgggttcccctatcacc
gtccaaccaaggggatagtaggtac
X232 ctagccggttgcttatcaggaatccc
ggccaacgaatagtccttaggggtac
X233 ctagcacggccagtaagtgtccaatc
gtgccggtcattcacaggttaggtac
X234 ctagcacacctgggcgtgaataactc
gtgtggacccgcacttattgaggtac
X235 ctagccttgggctcctgctaaaagac
ggaacccgaggacgattttctggtac
X236 ctagcgcgacctctaagccttgaaac
gcgctggagattcggaactttggtac
X237 ctagccaggtttgggggtacatgatc
ggtccaaacccccatgtactaggtac
X238 ctagctcgttgtctgccctcttgtag
gagcaacagacgggagaacatggtac
X239 ctagcagacctcggttaaccacgaaac
gtctggagccattgggtgctttggtac
X240 ctagcagtaaggctctttgggtgccac
gtcattccagaaacccacggtggtac
X241 ctagcagtagccttgtgggaaccaac
gtcatcggaacacccttggttggtac
X242 ctagcgattaacctctcggaatgcgc
gctaattggagagccttacgcggtac
X243 ctagcgcgatattgttccctgcttcc
gcgctataacaagggacgaagggtac
X244 ctagccaaccgtgttagcgaccatac
ggttggcacaatcgctggtatggtac
X245 ctagcgaatgggtggtggaaggatac
gcttaccacccaccttcctatggtac
X246 ctagcaccaagtcgctgtcaactgac
gtggttcagcgacagttgactggtac
X247 ctagccattttcaggaaggagacggc
ggtaaaagtccttcctctgccggtac
X248 ctagctggttggactgatacgaacgc
gaccaacctgactatgcttgcggtac
X249 ctagcattcacccaatgggtctacggc
gtaagtgggttaccagatgccggtac
X250 ctagcgggtggtaggcatatccgaaac
gccaccatccgtataggctttggtac

X251 ctagctgtctgtgaccttgggattgc
gacagacactggaaccctaacggtac
X252 ctagccaagcctcaatgcctaaaggc
ggttcggagttacggatttccggtac
X253 ctagcacttcttacgtcaccgcgatc
gtgaagaatgcagtggcgctaggtac
X254 ctagcggttcctcaccattcggttac
gccaaggagtggtaagccaatggtac
X255 ctagcagagtcaatgtcggttggctc
gtctcagttacagccaaccgaggtac

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☐ FADED TEXT OR DRAWING

☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.